AF MANUAL 52-31

guided missiles

FUNDAMENTALS

DEPARTMENT OF THE AIR FORCE

AF MANUAL 52-31

guided missiles

FUNDAMENTALS

DEPARTMENT OF THE AIR FORCE

AIR FORCE MANUAL
NUMBER 52-31*

DEPARTMENT OF THE AIR FORCE
WASHINGTON, D. C.
20 SEPTEMBER 1957

FOREWORD

This manual is a text for training personnel in the principles of guided missiles and a reference for all personnel concerned with guided missiles training and operation. It covers the operation, maintenance, and inspection of guided missiles control and guidance systems. It discusses electronics as applied to guided missiles, with the assumption that persons reading this manual will have a fundamental knowledge of electronics. This manual covers aerodynamics, propulsion, instrumentation, and electronic control and guidance systems involved in the guided missiles field.

Information contained in this manual is for training purposes and should not be followed if it conflicts with directive type information such as Technical Orders.

Recommendations or suggestions for the improvement of this manual are invited. Comments should be forwarded to the Director of Personnel Procurement and Training, DCS/P, Headquarters USAF, Washington 25, D. C.

BY ORDER OF THE SECRETARY OF THE AIR FORCE:

THOMAS D. WHITE
Chief of Staff

OFFICIAL:

J. L. TARR
Colonel, USAF
Air Adjutant General

DISTRIBUTION

Zone of Interior and Overseas

Headquarters USAF	300
Major Air Commands	10
Except: State Adjutants, ANG	1
Subordinate Air Commands	5
Wings	2
Special**	

*This manual is a conversion of Air Training Command Manual 52-12, August 1955. Existing copies of the Air Training Command Manual may be used.

**Commanders will requisition additional copies of this manual as required for individual issue to personnel holding AFS's 31150 and 31250 and to other personnel concerned with Guided Missiles. Commanders are authorized to requisition additional copies of this manual when required for training purposes.

Copies of this manual may be purchased through the Superintendent of Documents, Government Office, Washington 25, D. C.

Contents

Contents (continued)

Chapter 1

The Story of Guided Missiles

The science of warfare now has a new area to deal with — push-button warfare. Now it is possible to destroy a city by just pushing a button that would launch a bomb-carrying guided missile. This aircraft would have a nuclear warhead. In the future, our country's security may depend greatly on guided missiles of this type.

In the past, the Air Force has depended on the bomber for delivery of the atomic bomb. In using the bomber, the Air Force needed air superiority over the target and little or no hostile antiaircraft opposition in order to assure delivery of bombs.

But a guided missile can deliver the atomic bomb to its target without waiting for air superiority; it can get past any present-day means of interception. The guided missile can do this because it is capable of flying at supersonic speeds. Guided by its own electric brain, it is capable of destroying the target, whether it be a city factory, military installation, warship, or aircraft.

When an attempt is made at interception by conventional aircraft or guided missile, a missile may even be capable of detecting the interceptor's approach and firing smaller missiles in defense. The smaller missile, with its own electric brain and rocket engine, will seek out and destroy the oncoming aircraft or missile, leaving the bomb-carrying missile to continue its flight to the target undisturbed.

With the coming of guided missiles, a new career field came into being. This new field requires the services of thousands of skilled technicians and engineers, because the field involves the supersonic speeds and extremely high-altitude flying of missiles. The new speeds and new altitude outstripped and outpaced human ability to pilot aircraft. As a result, electronic or automatic missile guidance systems have had to be utilized.

The intent of this manual is to present to you the fundamentals of the missiles field — to inform you of the aerodynamics, pro-

pulsion, instrumentation, electronic control, and guidance systems involved in guided missiles. You will also learn of the capabilities of missiles, how they can be used, and what your job might be in a missile squadron.

What was once the briefing and ready room for pilots is now the control and guidance shop on a launching pad. On the launching pad the necessary adjustments are made on the missile equipment so that the missile will fly at the specified altitude and supersonic speed, arrive at a selected target, and destroy it.

The possibilities of guided missiles as an effective weapon are seemingly unlimited. Remember that the effectiveness of our missile attack is greatly dependent on you and the men with whom you work.

BRIEF HISTORY OF GUIDED MISSILES

Just as World War II was not fought along the same lines as World War I, victory in a future war will not be realized by World War II aircraft practices. However, the solution of guided missile problems of the future will stem from past experiences. If we are to profit from the lessons learned from the missiles of the past, these lessons must be analyzed and applied in the light of the current situation.

The idea of guided missiles was born during World War I. The use of the airplane as a military weapon brought about considerable thought concerning a remotely controlled aircraft which could be used to bomb a target. The leaders in the field were Orville Wright, who flew the first airplane; E. A. Sperry of the Sperry Gyroscope Company; and Charles F. Kettering of General Motors Corporation. It was these men who devised and tested our first missile, a small version of the aircraft used in those days. While the first missile did not get into combat, a most important result of these early tests was the recommendation that any future work should be done with radio-controlled aircraft, so that the missile could be given necessary adjustments while in flight.

In 1924, funds were allocated for developing a missile using radio control. Numerous, moderately successful flights were made during the 1920's with radio control. By 1932, however, the project had been listed in the files under frills and luxuries and closed because of lack of funds.

About 1935, two brothers named Good, amateur model airplane builders, built and flew a model plane that was remotely controlled by radio waves transmitted from the ground. These flights were the first completely radio-controlled flights on record.

Radio-controlled target planes were the first airborne remote-controlled aircraft used by our Army and Navy.

By December of 1941 just before our entry into World War II, remote-controlled aircraft were developed to the point where they were seriously considered for use as a weapon of warfare by General Arnold, then Chief of Staff of the Army Air Corps.

So far we have discussed only the missile powered by internal combustion engines and propellers. Work also was done to develop missiles using the reaction-type engines, including rocket engines, which contain within themselves all the elements needed for power, and jet engines, which depend on the surrounding atmosphere as a source of oxygen. When an atomic power plant for aircraft is developed, both the jet and rocket engines may become obsolete insofar as missile power plants are concerned. Let's now look at the progress rocketry has made in the United States.

Development of American Rocketry

Dr. Robert H. Goddard, at one time a physics professor at Clark University, Worcester, Massachusetts, was largely responsible for the sudden interest in rockets back in the twenties. When Dr. Goddard first started his experiments with rockets, no related technical information was available. He started a new science, industry, and field of engineering. Through his scientific experiments, he pointed the way to the development of rockets as we know them today. The Smithsonian Institute agreed to finance his experiments in 1920. From these experiments he wrote a paper titled "A Method of Reaching Extreme Altitudes," in which he outlined a space rocket of the step (multistage) prin-

ciple, theoretically capable of reaching the moon.

Goddard discovered that with a properly shaped, smooth, tapered nozzle he could increase the ejection velocity eight times with the same weight of fuel. This would not only drive a rocket eight times faster, but sixty-four times farther, according to his theory. Early in his experiments he found that solid-fuel rockets would not give him the high power or the duration of power needed for a dependable supersonic motor capable of extreme altitudes. On 16 March 1926, after many trials, Dr. Goddard successfully fired, for the first time in history, a liquid-fuel rocket into the air. It attained an altitude of 184 feet and a speed of 60 mph. This seems small as compared to present-day speeds and heights of missile flights, but instead of trying to achieve speed or altitude at this time, Dr. Goddard was trying to develop a dependable rocket motor.

Dr. Goddard later was the first to fire a rocket that reached a speed faster than the speed of sound. He was first to develop a gyroscopic steering apparatus for rockets. He was the first to use vanes in the jet stream for rocket stabilization during the initial phase of a rocket flight. And he was first to patent the idea of step rockets. After proving on paper and in actual test that a rocket can travel in a vacuum, he developed the mathematical theory of rocket propulsion and rocket flight, including basic designs for long-range rockets. All of this information was available to our military men before World War II, but evidently its immediate use did not seem applicable. Near the end of World War II we started intense work on rocket-powered guided missiles, using the experiments and developments of Dr. Goddard and the American Rocket Society.

The American Rocket Society was developing rockets and rocket motors after its organization in 1930. Its first motor was based mostly on German designs obtained from the German Rocket Society in 1931. The American Rocket Society was first to build a sectional rocket motor that could test motors of different sizes and shapes, thus cutting down the cost of a new motor for each type tested.

In 1941 some members of the American Rocket Society formed a company now known as Reaction Motors, Inc. It was organized to develop and manufacture rocket motors for both military and civilian use.

Development of German Rockets

The first flight of a liquid-fuel rocket in Europe occurred in Germany on 14 March 1931, five years after Dr. Goddard made his first successful test with liquid fuel. A German scientist named Winkler was in charge. Winkler lost his life a short time later during one of his experiments.

Germany by this time had begun to sense the future importance of liquid-fuel rockets in warfare. In 1932 General Walter Dornberger (then a captain) of the German army obtained the necessary approval to develop liquid-fuel rockets for war purposes. By 1936, Germany decided to make research and development of guided missiles a major project. Germany spent $40,000,000 on a project, known as the "Peenemunde Project," for establishing a large rocket research and development laboratory. Hitler put the members of the German Rocket Society to work there, closing to the rest of the world German developments on rockets until after the war Unlike Germany, the US during this time paid little attention to the development of jet and rocket propulsion for any specific purpose.

Evolution of Jet Engines

The rocket was just one type of jet propulsion power plants that was being proposed and worked on in this century. As early as 1913, Rene Lorin, a French engineer, proposed and first patented the idea for a ramjet power plant. Lorin's patent was followed by a Hungarian patent for a similar device in 1928 and another French patent in 1933. None of the proposed ideas resulted in a workable engine. The failures occurred not because the fundamentals of operating such a device were not known but because actual data on high-speed fluid flow was unavailable. A period of 32 years separated Lorin's original idea and the first free-flight testing of a ramjet powered vehicle capable of developing thrust in excess of drag. This test occurred in June

1945 when the Applied Physics Laboratory of John Hopkins University successfully flew the first ramjet powered aircraft.

The forerunner of the present day turbojet was not a thermaljet but a mechanical type. In 1927, the Italian Air Ministry began investigating the possibility of propelling an aircraft by placing a conventional type propeller inside the mouth of a venturi-shaped fuselage. This so-called "ducted propeller" installation was a form of mechanical jet propulsion. Tests with this "ducted propeller" installation demonstrated that it possessed excellent maneuverability and stability characteristics, although its overall performance was only mediocre. In 1932, an Italian by the name of Campini designed and later flew an aircraft propelled by a thermaljet engine. His jet-powered aircraft was not, however, a turbojet, because it depended upon a conventional reciprocating engine instead of a gas turbine for compressor power.

Evaluation of Campini's engine had hardly been made when improved jet engines began to appear in various countries. A young British engineer and Royal Air Force officer, Frank Whittle, had filed a patent for a thermaljet engine as early as 1930. Whittle's design eliminated the reciprocating engine as the power source for driving the compressor. Instead, the mixture of air and gases was used after combustion to drive a gas turbine. The turbine drove the compressor. On 7 April 1941, a Gloster "Pioneer" aircraft, powered by Whittle's engine, became airborne during taxiing tests and flew about 150 yards at an altitude of about 6 feet. On 15 May 1941, this same aircraft made the first official takeoff for a turbojet powered aircraft and flew for 17 minutes.

After these successful flights, the US Air Corps sent a special group of men to England to study the engine. Further development became a British-American project. At this time, only 10 hours of jet engine operation had been accumulated in fifteen flights.

Turbojet development in the US was turned over to the General Electric Company, because of its experience in the development and production of turbosuperchargers. Today, turbojet engines are built and developed by nearly all aircraft engine companies. At the present time, the turbojet engine is the only air-jet engine capable of getting an aircraft or missile to take off under its own power without the assistance of a booster.

Another air-jet engine is the pulsejet. This type of engine was patented by a German engineer in 1930, but a good, workable pulsejet engine was not perfected until World War II. It became famous during the war as the power plant of the German buzz bomb or V-1. This engine was capable of propelling the V-1 at 450 miles per hour.

In 1940, General Motors Corporation was given a contract to build and develop jet-powered, controllable bombs which in the final version were to be command-controlled with television. These bombs were tested extensively in 1941 and 1942. The testing led to the development of new types of jet-powered bombs.

Enemy Guided Missiles of World War II

The Japanese were far behind the Germans in developing missiles during World War II. The baka bomb used by the Japanese during the war was not a guided missile in the true sense of the word. It was a rocket-propelled, piloted glide bomb designed for use against shipping targets. The baka bomb was known as a suicide bomb. Although it achieved a certain degree of success, it had poor maneuverability, a characteristic which resulted in many of them being shot down by antiaircraft fire.

The Japanese also tried an air-launched, radio-controlled, rocket-assisted glide bomb. This missile had to be dropped from a low altitude, and the control plane had to get within 2½ miles of the target. The procedure made the control plane an easy target for antiaircraft fire. The project was dropped before the end of the war.

The German developments in the field of guided missiles during World War II were the most advanced. Their most widely known missiles were the V-1 and V-2 surface-to-surface missiles. As early as the spring of 1942, the original V-1 had been developed and flight tested at Peenemunde. In 1943, Germany was working on 48 different antiaircraft missiles. These were later consolidated into 12 projects

for immediate development into useful weapons. Toward the end of the war, all efforts were being directed toward the successful production of an antiaircraft missile capable of intercepting allied bombers.

The V-1 (a robot bomb) was a pilotless, pulsejet, midwing monoplane, lacking ailerons but using conventional airframe and tail construction. All guidance and control was accomplished internally by gyro stabilization and present compass guidance. It was launched from a ramp 150 feet long and 16 feet above the ground at the highest end. A speed of approximately 200 mph had to be reached before the V-1 propulsion unit could maintain the missile in flight. The missile carried a warhead weighing 1988 pounds.

The V-1 (vengeance weapon #1) was not accurate, and it was susceptible to destruction by antiaircraft fire and aircraft. However, the interruptions it caused in the functioning of a vital war center such as London, together with the amount of physical damage it did, made the V-1 effective in lowering morale.

The V-2 (vengeance weapon #2) was the first long-range, rocket-propelled missile to be put into combat. Concentrated efforts began in 1941. The V-2 was put into mass production and the first V-2 landed in England in September 1944.

The V-2 was a supersonic missile, launched vertically and automatically tilted over a 41- to 47-degree angle a short time after launching. The maximum range was about 200 miles, and the top speed was about 3300 mph. The V-2 was a large missile, having a length of 46 feet 11 inches and a diameter of 5 feet 5 inches. Its total weight at takeoff was over 14 tons, including a 1650-pound warhead.

Active countermeasures against the V-2 were impossible. Except for its initial programmed turn, it operated as a free projectile at extremely high velocity.

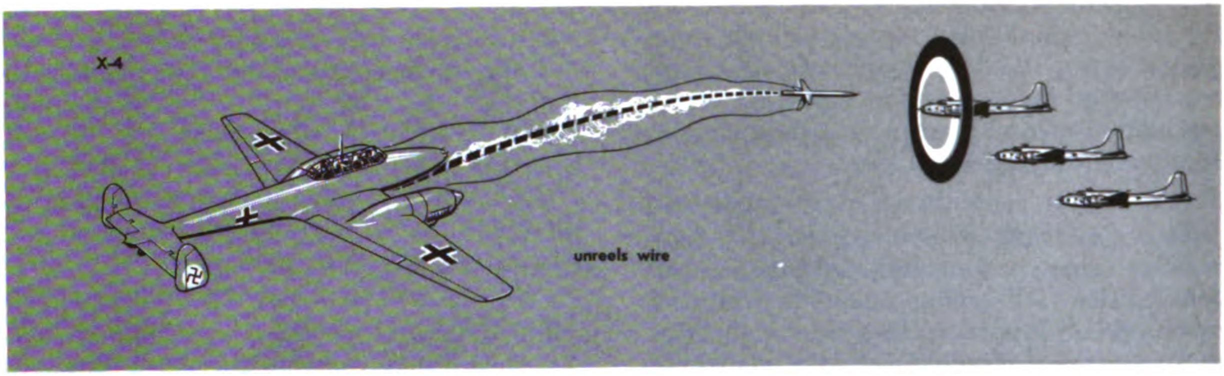
Five other German missiles were also highly developed during World War II and were in various stages of test.

One of these, the Rheinbote, was also a surface-to-surface type missile. This rocket was a three-stage device with booster-assisted takeoff. Its range was about 135 miles, with the third stage reaching over 3200 mph in about 25 seconds after launching. Overall length of the rocket was about 37 feet; but after having dropped a rearward section at the end of both the first and second stages, it had a length of only 13 feet. The 13-foot section of the third stage carried an 88-pound high-explosive warhead.

A surface-to-air missile, the Wasserfall, was a radio-controlled supersonic rocket, similar to the V-2 in general principles of operation. Fully loaded it had a weight of slightly less than 4 tons. Its length was 25 feet. Designed for intercepting aircraft, this missile had specifications which called for maximum altitude of 65,000 feet, speed of 560 mph, and range of 30 miles. Its 200-pound warhead was to be detonated by radio after the missile had been command-controlled to its target by radio signals. It also was to use an infra-red proximity fuse and homing device for control on final approach to the target and for detonating the warhead at the most advantageous point in the approach. Propulsion was to be obtained from a liquid-fuel power plant, with nitrogen-pressurized fuel tanks.

Another surface-to-air missile, the Schmetterling (HS-117), was still in the development stage at the close of the war. All-metal in construction, it was 13 feet long and had a wing span of 6½ feet. Effective range against low-altitude targets was 10 miles. It traveled at 540 mph at altitudes up to 35,000 feet. It was to use a proximity fuse to set off its 55-pound warhead. Propulsion was obtained from a liquid-fuel rocket motor with additional help from two booster rockets during takeoff. Launching was to be accomplished from a platform which could be inclined and rotated toward the target.

A third German surface-to-air missile was the Enzian, designed to carry pay loads of explosives up to 1000 pounds. It was to be used against heavy-bomber formations. The Enzian was about 12 feet long. It had a wing span of approximately 14 feet and weighed a little over two tons fully loaded. Propelled by a liquid-fuel rocket, it was assisted during takeoff by four solid-fuel rocket boosters.



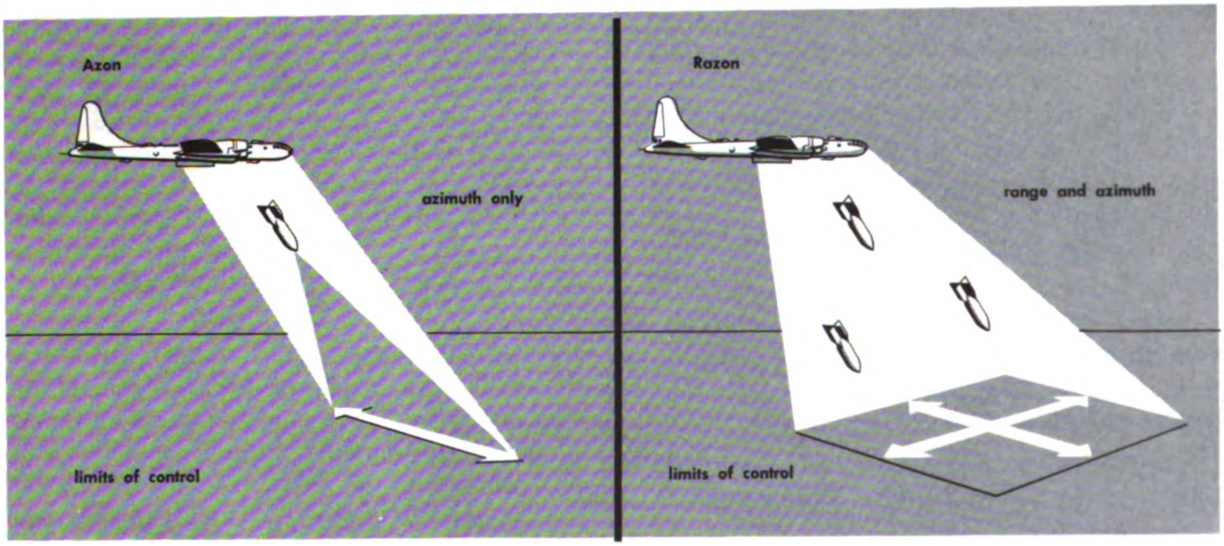
Launching of an X-4

Launching, as in the case of the Schmetterling, was from a rotatable launcher, with range elevation possible. Its range was about 16 miles, speed 560 mph, and maximum altitude 48,000 feet. Guidance was by command control. It was believed to be gyroscopically stabilized in roll.

A German air-to-air missile, the X-4, was designed to be launched from fighter aircraft as shown in the illustration at the top of this page. Propelled by a liquid-fuel rocket, it was stabilized by four fins placed symmetrically. Length was about 6½ feet and span about 2½ feet. Its range was slightly over 1½ miles, and its speed was 560 mph at an altitude of 21,000 feet. Guidance was accomplished by electrical impulses transmitted through a pair of fine wires from the fighter aircraft. The wires unrolled from two coils mounted on the tips of two opposite fins of the missile. This missile was claimed to have been flown, but it was never used in combat.

United States Guided Missiles of World War II

A project for developing missiles in the US during World War II was instigated in 1941. In that year the Army Air Corps asked the National Defense Research Committee to undertake a project for the development of a vertical, controllable bomb. The committee initiated a glide-bomb program which resulted in standardization of a preset glide bomb attached to a 2000-pound demolition bomb.



Limits of control of the Azon and Razon

The Azon, a vertical bomb controlled in azimuth only, went on the production line in 1943. Project Razon, a bomb controlled in both azimuth and range, was started in 1942 but not completed until the end of the war. The limits of control of both Azon and Razon bombs are shown at the bottom of page 6. A medium-angle glide bomb called the ROC and a 12,000-pound bomb known as the Tarzon, both controllable in azimuth and range, were also under development at this time. The two bombs did not reach the combat stage during World War II. The tarzon project was dropped in 1946 and picked up again in 1948. The Tarzon was used successfully in the Korean action.

In 1943, a project was initiated for development of a glide torpedo. Standard Navy torpedoes were used for this project. In the final days of the war, these glide torpedoes were used on several missions in the Pacific theater.

In 1944, we carried out a glide-bomb mission against Cologne, Germany. A majority of the bombs reached the target area. In this same year remote television-control equipment was developed and installed in bombing aircraft. These aircraft were used to control television-sighted, explosive-laden bombers unfit for further service. These radio-controlled bombers saw some service over Germany under the title of the "Weary Willie" project. Light and radar target-seeking devices were developed for use with glide bombs and were tested until 1945.

Our first jet-propelled bomb was a flying-wing, jet-powered, and radio-controlled. The second version of a jet-propelled bomb was a copy of the German buzz bomb with a few improvements. Another model consisted of a combination of the two mentioned above, using the flying wing together with the pulse-jet engine of the buzz bomb. This project become obsolete in 1946. By 1946 the Tiamat, a guided aircraft rocket, had been completed to the extent that full-size versions were being tested.

The Navy had a number of guided missile projects under development by the end of the war. One of these, the Gargoyle, was an air-launched, powered, radio-controlled glide bomb with a flare for visual tracking. The Navy also had a glide bomb called the Glomb. It was guided to a target by radio control, monitored by television. The Loon, a modification of the German V-1, was to be used from ship-to-shore and to test guided missile components. Another missile, known as Gorgon IIC, used a ramjet engine with radar tracking and radio control.

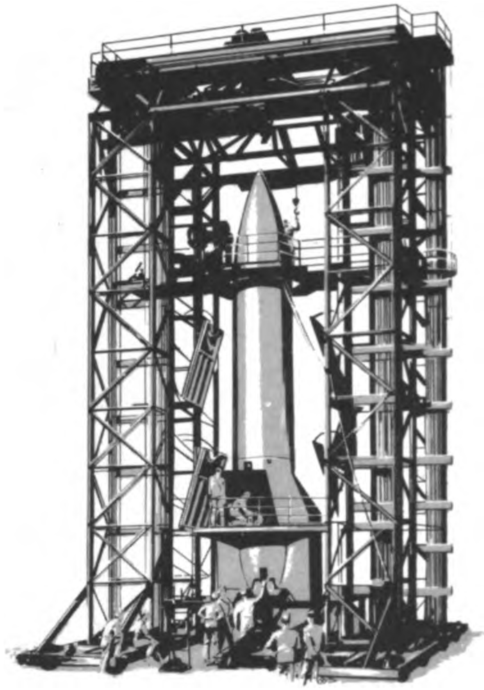
During the war these weapons were developed under pressure for immediate use. At the end of the war in 1945, nearly all previous development on guided missiles, controllable bombs, and guided aircraft rockets was considered obsolete. New military characteristics and specifications were drawn up with future weapon possibilities in mind.

GUIDED MISSILE DEVELOPMENT AND CLASSIFICATION

A high respect for the far-reaching aspects of guided missiles was conceived in the closing phases of World War II. Immediately following the war's end, the Army Air Corps began a crash program of missile development. Air Corps policy in the development of guided missiles, as stated in February 1946, was:

> . . . to limit the improvement or modification of existing guided missiles to the incorporation of such new discoveries as will greatly improve their characteristics, or the incorporation of such elements as appear to have definite application to the guided missiles of the future. With the exception of such extremely limited improvement as may prove desirable in existing guided missiles, AAF missile research and development is being directed towards discovering the many unknowns, in order that an intelligent decision as to the logical approaches to, and the type of missile weapon required for, the missile force of the future may be missile made.

This program was later drastically curtailed because of peacetime budgeting and the need for conservation of limited numbers of



Readying a missile for launching

technically qualified personnel and limited supplies of materiel. Contracts were cancelled for interim-type guided missiles which were being developed as so-called "insurance" until such time that longer range and supersonic missiles could be realized.

Much effort is presently being concentrated on studies of supersonic speed phenomena; upper atmosphere conditions; and basic problems connected with propulsion, control, electronics, guidance, and launching. Now under development or already developed for future employment are various surface-to-air, air-to-surface, air-to-air, surface-to-surface, and air-to-underwater guided missiles. Some examples of such missiles are shown in the illustrations on accompanying pages.

A research and development program starts with the need for a missile to accomplish a particular type of mission. The size, range, speed, and accuracy are specified. In each of these missiles, there are certain characteristics which have priority from a military viewpoint. Accuracy and speed are two characteristics which rank high on all priority lists. No missile is operationally successful unless we can place it accurately enough on the target to justify the expenditure of money, manpower, and materiel. Speed sufficiently high to avoid detection, or at least to gain appreciable immunity from countermeasures, is also important.

Reliability of guided missile operation can be developed and maintained only by the process of exposing the missile's many components to an exhaustive testing program. Also, margins of safety between service and working conditions must be established. The extreme conditions under which individual components fail must in like manner be determined.

The program of component research is based on realizing the major aims and overcoming the problems which have been plaguing the development of guided missiles from the beginning. Research has determined, for example, that liquid rockets are the best propulsion devices for high-altitude supersonic missiles, solid-fuel units are superior for booster charges, gas turbines and other jet engines are more economical in operation but intricate and costly to manufacture, and ramjets hold great promise for altitudes up to 70,000 feet. Launching equipment is concerned with the problem of accelerating supersonic missiles to proper velocity, a problem which usually involves boosters weighing many tons.

How a Piece of Equipment Evolves

Before seeing how a piece of equipment evolves, let's look at the organizational make-up of the agencies involved in research and development of a piece of equipment.

RESEARCH AND DEVELOPMENT ORGANIZATION. The chart on the following page illustrates the relationships existing between all agencies of research and development as to staff and command functions. The heavy black connecting lines indicate direct chain of command, starting with the Secretary of Defense and working to the right, to the various commands. Lighter connecting lines indicate staff personnel and staff agencies to the various organization heads.

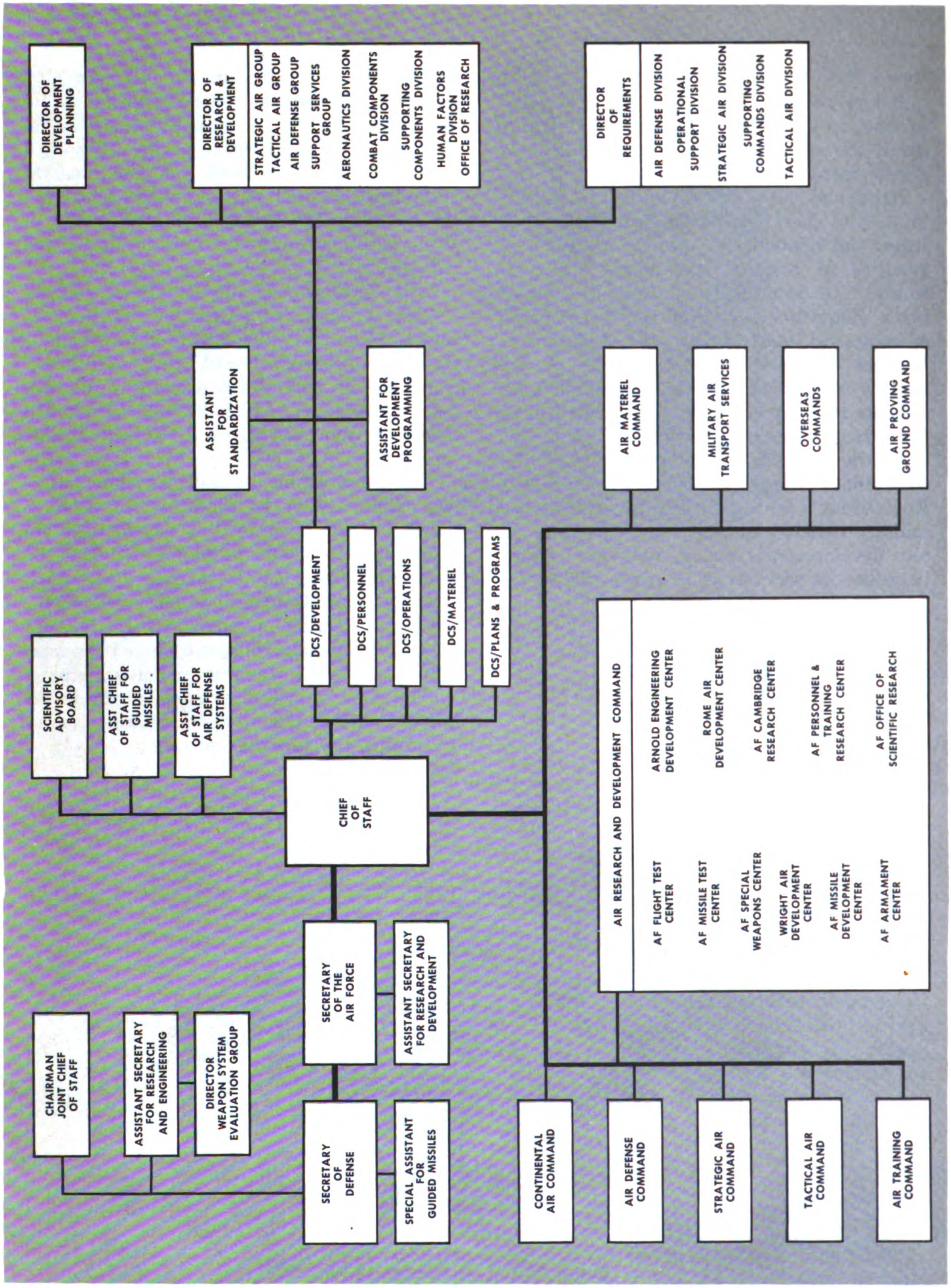
At the Department of Defense level, the Assistant Secretary of Defense (Research and Engineering) is responsible for the entire military research and development effort. He sets developmental policy, resolves disputes between the services as to responsibility for specific developments and handles the allocation of research and development funds. He is assisted by the Weapons Systems Evaluation Group (WSEG) which works closely with the Joint Chiefs of Staff and a number of technical Coordinating Committees which act in an advisory capacity. The Secretary of Defense also has a Special Assistant for Guided Missiles.

The Secretary of the Air Force has an Assistant Secretary for Research and Development. Note that the Chief of Staff has a Scientific Advisory Board (SAB) which advises him on technical matters of all natures and an Assistant Chief of Staff for Guided Missiles who specializes in guided missile programs.

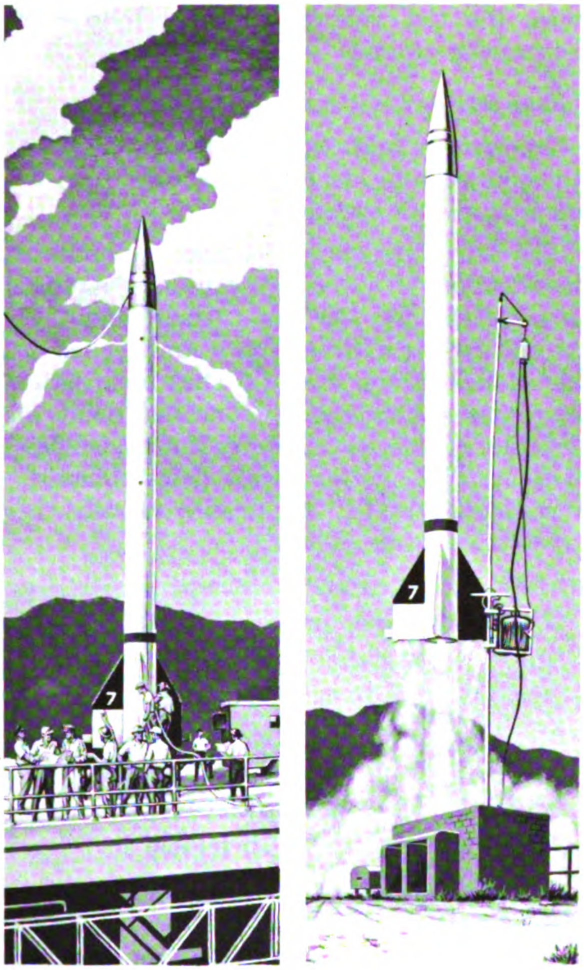
The Deputy Chief of Staff, Development, is the chief advisor to the Chief of Staff in matters pertaining to the development of equipment. It should be emphasized that the interests of the DCS/Development are much broader than straight materiel development. DCS/Development is charged with development organization, budgetary methods, personnel utilization, planning, and programming.

The Assistant for Development Programming handles budget, financial, and facility matters in the development area.

TM-61 Matador on launcher . . . US Air Force



Research and Development Organization of the USAF



Viking—US Navy

The Director of Development Planning is concerned with the future and establishes the framework of the long-range research and development program.

The Air Force Director of Research and Development (AFDRD) is the Air Staff agency with most direct supervision of the Air Research and Development Command (ARDC). This staff agency has groups responsible for weapon systems in the strategic, tactical, air defense, and supporting services areas and divisions and branches corresponding to the major technical areas of Air Force research and development as well as a Human Factors Division and an office of Research which is concerned with basic scientific research.

The Air Force Director of Requirements (AFDRQ) is organized on a mission basis. As shown in the organization chart, this agency includes the tactical, strategic, air defense, operational support, and supporting commands divisions. AFDRQ establishes military requirements for new equipment, new methods, new techniques, etc, and transmits them to the Director of Research and Development for the action necessary to fulfill them. The Director of Requirements also has supervision of the Air Proving Ground Command, where new items are tested to determine whether they meet operational requirements.

The Air Research and Development Command (ARDC) is a major operating command taking its instructions from the Chief of Staff, USAF, usually through DCS/Development or AFDRD. You can see that under ARDC are a number of centers, such as the flight Test Center and Wright Air Development Center, where the actual materiel development work is done. ARDC also lets development contracts to industry.

DEVELOPING A PIECE OF EQUIPMENT. The development process usually begins in the field when an operational command needs a new piece of equipment. This requirement is transmitted to the Director of Requirements who validates it, draws up a formal operational requirement which is transmitted to the Director of Research and Development. AFDRD determines the best course of action to meet this requirement and fits it into the overall picture. The director then sends a development directive to ARDC.

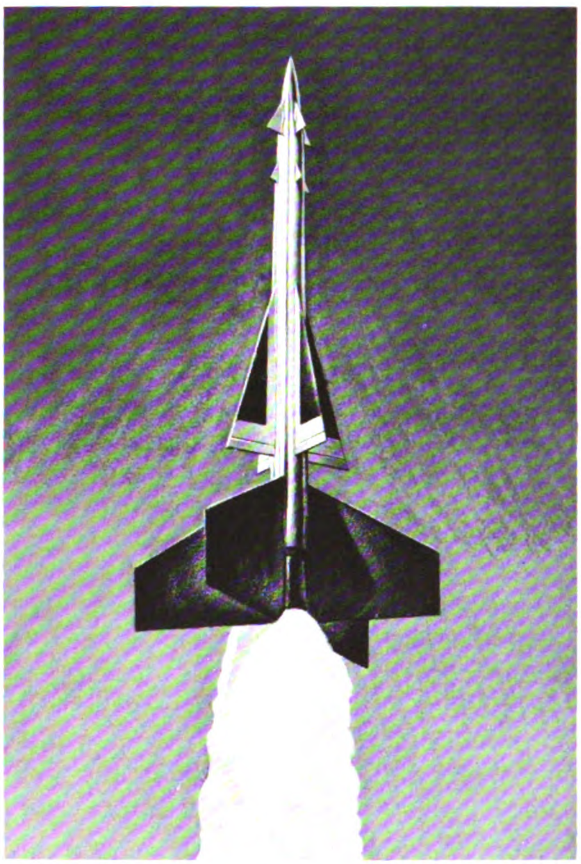
In the case of major pieces of equipment, AFDRD and ARDC work on the basis of a complete weapons system. ARDC tries to keep this complete package in view so that a complete operational system can be delivered at the required time. ARDC begins testing when it believes it has something which meets the requirement. When satisfied, it turns over test production to Requirements. The Operational Test Center conducts tests at Eglin AFB to see whether the system meets the operational requirement. If it does, it is standardized and put into production. The finished equipment goes out to using commands, thus completing the cycle.

The agencies connected with overall development in the Air Force are shown in the

chart at right. Heavy lines indicate the direct line of control. Lighter lines show further division of control to the individual agencies of the Air Force which do research and evaluation.

Notice that the chart shows the hub of the network of agencies to be the Deputy Chief of Staff Development.

The three phases of a missile testing program are: preflight testing, consisting of a series of tests upon each component and on the assembled missile; flight testing, involving the checking of launching problems, aerodynamic stability, flight characteristics, performance under guidance, and performance against the target; and the third phase, service testing of the missile by armed forces units. Work on components is usually done through the aircraft industry by its subcontracting to electronics companies, research laboratories, and universities. When components have been assembled for test under actual flying conditions, the test vehicle is moved to one of the proving grounds.



Nike Missile US Army

Guided Missile Classification

To study missiles in a systematic manner, it is necessary to identify or classify them by some simplified system.

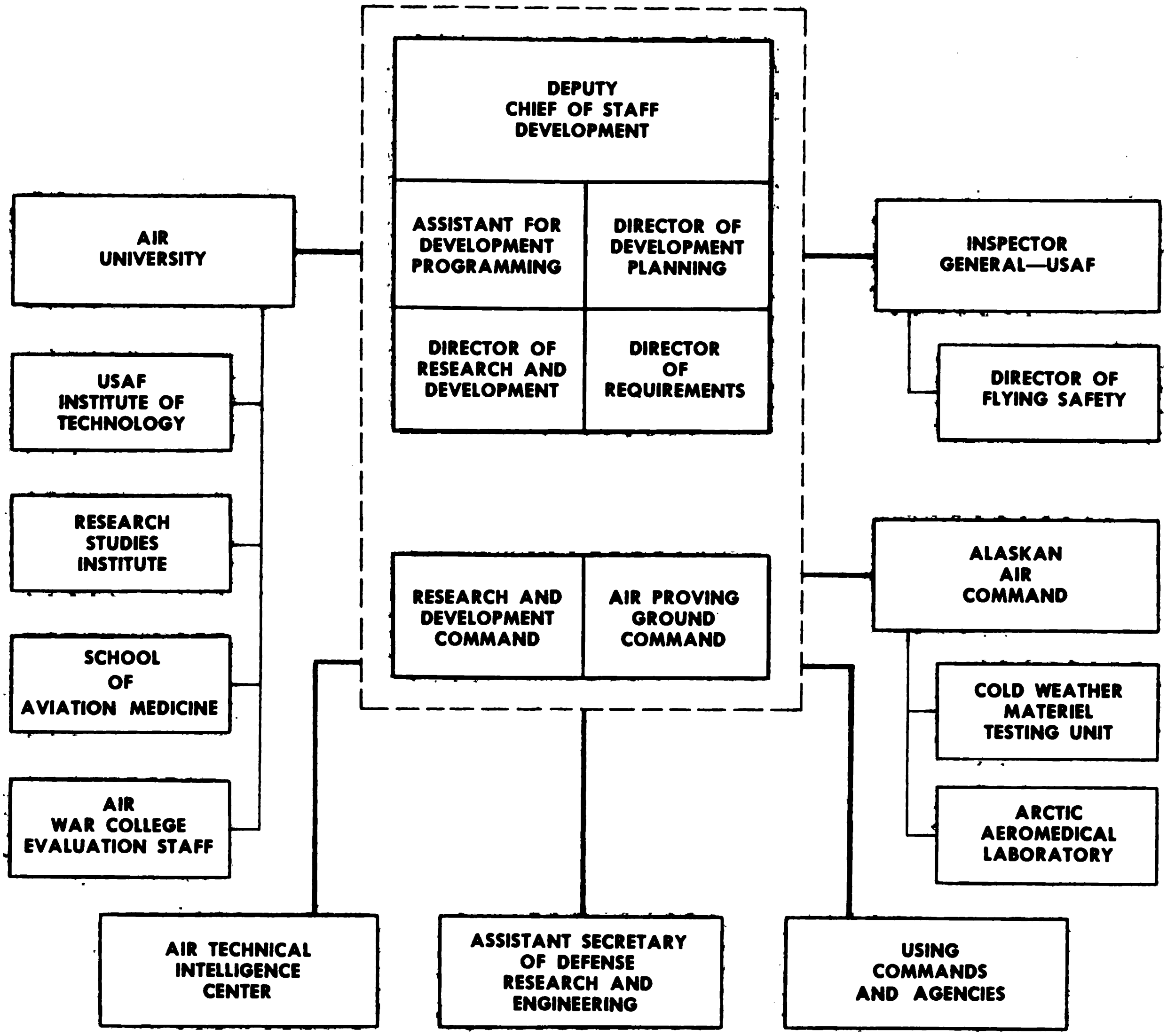
The Air Force employs two methods of classification. The most common method uses "A" for air, "S" for surface, and "U" for underwater. These letters are applied in the classifications below. The first letter designates the intended launching point of a missile, and the second letter is intended destination. These letters, used in combination with the letter "M" for missile, classify the use of missiles as follows:

SAM—Surface-to-air missile
AAM—Air-to-air missile
ASM—Air-to-surface missile
SSM—Surface-to-surface missile
AUM—Air-to-underwater missile
SUM—Surface-to-underwater missile
USM—Underwater-to-surface missile
UAM—Underwater-to-air missile

In this method of classification, missiles used by any one branch of the service are so identified by a service letter immediately following the classification and separated from it by a hyphen. The service letter "A" designates the Air Force, "N" the Navy, and "G" the Army. When a missile is used jointly by more than one service, the service letter of each using agency is shown. For example, if all three services are involved, the letters "ANG" are used.

Models are numbered beginning with "1." The model number may be followed by a letter (not capitalized) designating each succeeding modification. Model designation follows the branch designation as follows: ASM-G-3c, the 3c being the model designation.

Status of a missile when under research and development, service test, actual service, or retirement is indicated as follows: "X" means experimental, "Y" service test, and "Z" obsolete. These letters, when used, precede the basic classification. No prefix indicates an accepted service missile.



Agencies concerned with the overall development of the Air Force

When a missile is used as a test vehicle, the letters "TV" followed by the service letter, model number, and modification letter are its means of identification. The type of testing is indicated by the letter preceding the test-vehicle designation as follows:

A—Aerodynamics

C—Control

L—Launching

P—Propulsion

R—Research (includes high-altitude sounding rockets)

Conventional aircraft are sometimes employed as missiles for the purpose of testing missile systems. In such cases the standard aircraft designation is prefixed by the letter "M" to indicate missile aircraft.

When conventional aircraft are modified to serve as controlling or directing aircraft for guided missiles, the standard or basic aircraft designation is prefixed by the letter "D" to indicate director aircraft. Here are several examples of classification for your consideration:

ASM-G-3c—Air-to-surface missile used by the Army, third model and third modification, accepted service.

XSSM-A-lb — Experimental surface-to-surface missile used by the Air Force, first model, second modification.

YAAM-N-2d — Service test air-to-air missile used by the Navy, second model, fourth modification.

Lark CTV-N-9b — Lark (missile name), control test vehicle used by the Navy, ninth model, second modification.

DF-80 — F-80 fighter used as a director aircraft (commonly called a "chase plane").

The most recent form of classification adopted by the Air Force has the following designations: TM-61, SM-62, SM-64, SM-65, IM-99, GAR-1, and GAM-63. TM means *technical missile;* SM, *strategic missile;* IM, *interceptor missile;* GAR, *guided aircraft rocket;* and GAM, *guided aircraft missile.*

Another form of classification is that contained in AFL 136-3, 18 September 1952, "Guided Missiles in the United States Air Force." This classification divides guided missiles, in a rather broad sense, into three types: the B-61 bomber type (SSM), F-99 fighter type (SAM), and guided aircraft rockets such as the GAR-1 (AAM).

The practice of giving a popular name to different types of aircraft, in addition to their technical designation, has been in existence for many years in the Air Force. The following popular name groups are authorized for missiles.

AAM — Winged creatures (excluding birds of prey and game birds)
ASM — Birds of prey
SAM — Mythological terms
SSM — Astronomical terms
Target drones — Game birds or hunting terms

Exceptions to this type of grouping are missiles named prior to the use of this system; they retain their original names. An example of such an exception is the Lark.

Guided missiles also include research drones and target drones. Research drones are generally conventional aircraft remotely controlled from a mother ship. Target drones are of conventional design with the exception of their size. A typical target drone is the OQ-19, which has a wing span of about 12 feet. This type of target drone is used for training aircraft gunnery crews.

While the following methods of classification are not commonly used, they do offer further means of separating types of missiles. Briefly, missiles can also be classified in terms of their propulsion, as to air-independent (rockets), or air-dependent (air-jet engines).

Missiles also are classified as to Mach number, which compares the missile's velocity to that of the speed of sound.

Subsonic, sonic, and supersonic are terms which also relate the speed of a missile to that of sound.

Classification as to the type of guidance system used is sometimes employed. In this method, the missiles are classified as to whether the terminal type of guidance system is active, semiactive, or passive. These systems all use the detection of some radiation as a means of guiding the missile to the target. The missile may detect radiation automatically given off by the target (passive type), or it may detect reflections of radar signals transmitted either by the missile itself (active type) by ground-based equipment (semiactive type).

Guided missiles may also be classified in terms of their range, *long* or *short,* the present limit of short range being 500 miles.

You will find that to "talk shop" effectively and easily you need to know the classifications of missiles. By knowing them you'll know which type of missile an airman is talking about when he refers to a XSSM-A-lb.

Missiles Have Come a Long Way

The statement that guided missiles have shown considerable progress from World War I days to the present time remains valid despite the many years missile development lay in a dormant state. The progress made can be attributed mostly to the independent research and development accomplished in such fields as electronics, rocketry, jet propulsion, and aerodynamics. The future outlook for missiles, although presenting many obstacles to be overcome, is that they will be the main weapon of war.

Chapter 2

Aerodynamics of Guided Missiles

The detailed study of air in motion and the mathematical analysis of various forces present are involved and lengthy. We will leave such detailed study for the aircraft engineer and confine our discussion to the basic theories and principles involved in missile aerodynamics.

PHYSICS OF FLIGHT

In a study of aerodynamics, you need to understand several basic laws of physics. Once these laws are understood, applying them to the particular problems of missile aerodynamics becomes relatively easy.

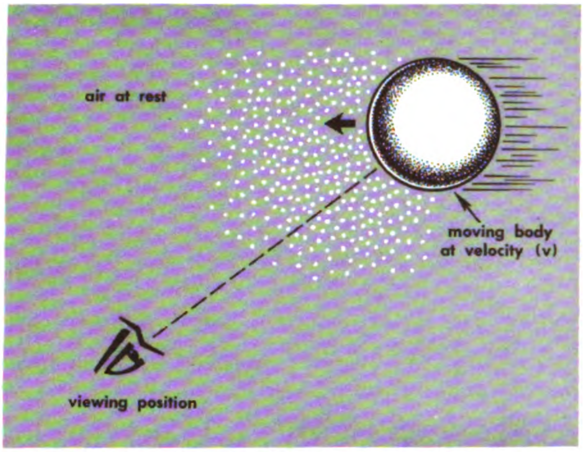
In general, missile aerodynamics are the same for subsonic and supersonic flight. We can go even further and say that, in general, the aerodynamics are the same for any craft that is intended to fly; that is, in order to fly, the craft must be aerodynamically sound. However, because of our rapid advancement in speeds and altitudes, there are new concepts and problems that arise, such as the problem of the so-called shock wave which occurs when the speed of sound (sonic speed) is reached. Problems regarding oxygen and heating arise at high altitudes.

First, we will give consideration to phenomena that are common to both subsonic and supersonic speeds. The problems of high-altitude flying will be taken up as we meet them.

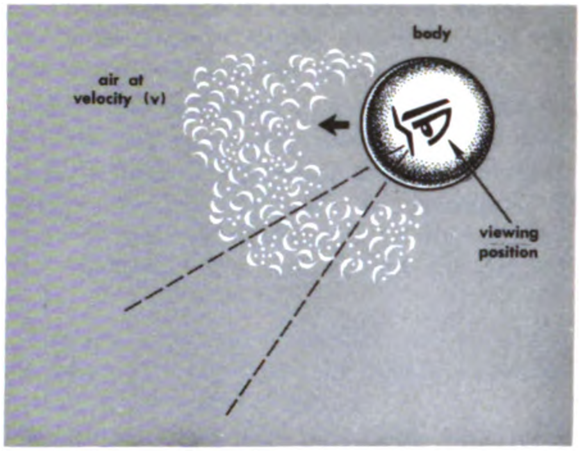
Relativity of Motion

To start with, it will be helpful to consider some of the forces which act on a body that is moving through the air. If you were to watch a body moving through the air, it would appear as though the air were standing still. This is the impression you get when watching an airplane fly overhead. In this case it would seem that the opposing force on the body would be due to the velocity of the body moving through the air as shown in the first illustration on the next page.

Now if you were riding on the body, it would seem that you were standing still and



Observations from a stationary point



Observations from a moving body

the mass of air was moving past you. This situation is represented in the next figure.

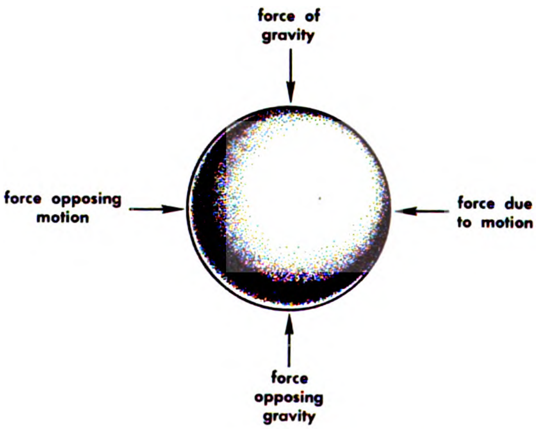
In the preceding examples you have the basic concept of *relativity of motion*. The forces exerted on the body by the air, in either case, are the same. This principle may be stated as: *The force exerted by the air on a body does not depend on the absolute velocity of either, but only on the relative velocities between them.*

In the study of aerodynamics, we put the principle just stated into practice by the use of wind tunnels in which the aircraft are held stationary and high wind velocities are made to pass over them. This simplifies the study of forces acting on the airframe and enables us to virtually flight test missiles while they are standing still.

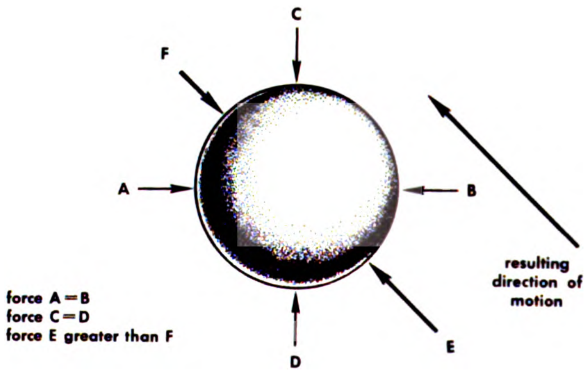
Forces Acting on a Missile

There are numerous forces acting on all parts of a missile, which are due to *air resistance*, *gravity*, *friction*, and/or other factors. Consider again the body moving through the air at a certain speed. There are four general forces present. One is the force the body exerts on the air in moving forward. In opposition to this is the force the air delivers to the body. The other two forces are — the force present because of gravity forcing the body toward the earth and the force the body exerts in the opposite direction to keep itself aloft.

In the left illustration below, the body would not move if all the opposing forces were equal; instead, the body would remain in a state of equilibrium. If any of the forces acting



Forces acting on a body moving through air



Unequal forces acting on a body

on the body were not equal, a movement would result in the direction of the greater force. For example, in the preceding illustration you see a representation of the forces acting on a body. The length of the arrows show the respective magnitudes of the forces, and the arrowheads point in the direction of the forces. The illustration shows that force "A" is equal and in the opposite direction to force "B," force "C" is equal and opposite to force "D," and force "E" is *greater* than its opposing force "F." As indicated, the body would move in the direction of the greater force "E."

The preceding figure exemplifies what is known as a *vector representation* of the forces acting on a body. Any number of forces may be shown by a vector representation and can be resolved, or simplified, into resultant forces to find what effect may result. Force diagrams are essential in designing structures that are to be subjected to a number of different forces.

NEWTON'S LAWS OF MOTION. Now let us examine three basic laws of physics known as *Newton's laws of motion*.

The first law states: "A body in a state of rest remains at rest, and a body in motion tends to remain in uniform motion unless acted upon by some outside force." This law tells us that whenever there are unequal forces upon a body, the body must move. After we have that body in motion it will stay in motion as long as there is no force present to *change* that motion. For example, if you were to push against a book that was lying on a table, you would find that there was a certain amount of force required to overcome *friction* to set the book in motion. If you could eliminate all of the restraining forces acting on the book once it is in motion, the book would continue to move uniformly until acted upon by some outside force. It is these restraining forces with which we are mainly concerned in the study of aerodynamics.

Newton's second law states: "The rate of change in the momentum of a body is proportional to the force acting on the body and is in the direction of the force." The momentum of a body may be defined as the force one body would exert in resisting any change of its motion.

His third law states: "To every action there is an equal and opposite reaction." This law tells us that if a force is applied, there must be a reaction opposite to and equal to the applied force.

DEFINITION OF FORCE. All along we have spoken of *force*. What do we mean when we say there is a "force applied"? Let's consider a pound of air moving in some direction. The pound of air is capable of producing a force. The force of that pound of air is *directly* proportional and equal to the weight and the velocity change. This can be stated simply as:

Force = Mass times Acceleration

or

$$F = ma$$

Since air has mass, once we put air into motion it is capable of applying a force.

Whenever a force is applied through some distance, we say it has done work. In simple form:

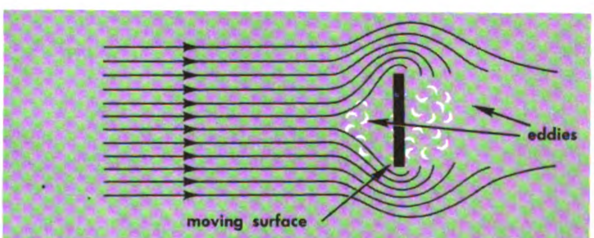
Work = Force times Distance

or

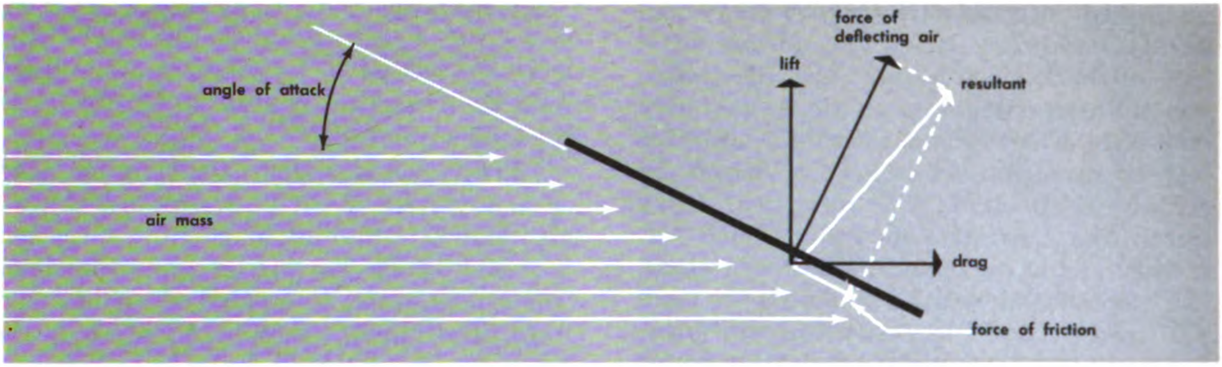
$$W = Fd$$

Any mass that is in motion is capable of applying a force and doing work. Whenever the motion of a mass is changed, there is, as stated earlier, a change in momentum. A change in the momentum of air sets up a turbulent reaction in the air. For our purposes, we will analyze the effects of a change in the momentum of the air by studying the various surface configurations and airflow paths.

EDDIES. When a mass of air moves over a surface there are a certain number of particles of the air that tend to rebound or swirl past the surface as shown below. The turbu-



Eddies resulting from air flow over a flat surface that is at a right angle to the air flow



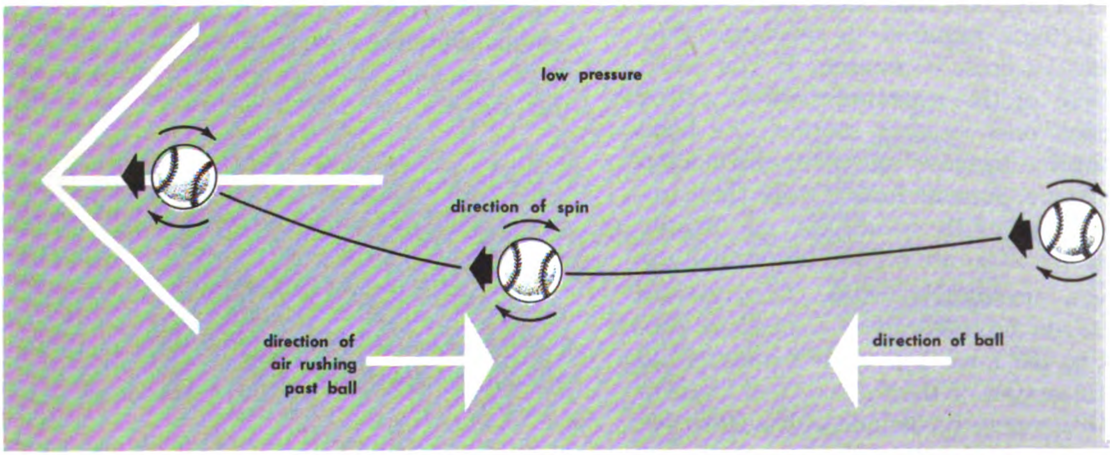
Forces acting on a flat surface in an airstream

lence that is set up by the swirling particles is known as eddies. The number of eddies depends on such factors as smoothness, size, and shape of the surface. Their number also varies with the angle the surface makes with the mass of air, with relative speeds, and with the density of the air.

LIFT AND DRAG. Because of the *friction* between the mass of air and the inclined surface as illustrated above, there is a force which tends to move the surface *parallel* to itself. There is also a force perpendicular to the surface, caused by the deflecting airstream. These combined forces give a resultant force backward in the direction of the moving air mass. Breaking this force down, we have one force perpendicular to the moving mass of air called lift and another force parallel to the mass of air called drag. The angle that the mass of air makes with the surface is called the *angle of attack*. The figure shows the forces which make up the *resultant* force on the surface at a particular angle of attack.

Bernoulli's Theorem. Scientist Daniel Bernoulli discovered that the total energy in any system remains constant. In other words, if one element of any energy system is increased, another decreases to counterbalance it.

Before applying Bernoulli's theorem to a wing section, let's consider what happens to a baseball when a pitcher throws a curve. By snapping his wrist, the pitcher puts a spin on the ball. Thus, after leaving the pitcher's fingers, the ball has two motions as shown in the following diagram. The ball is turning on an axis perpendicular to the ground, and it is moving forward.



Bernoulli's theorem applied to a pitched baseball

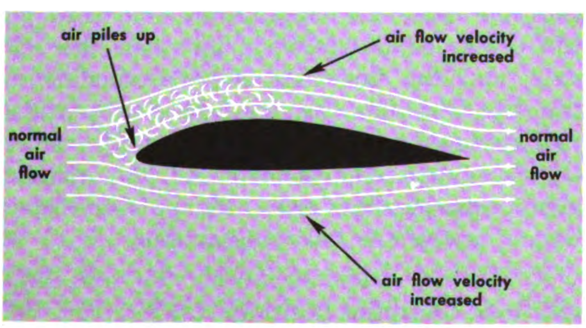
These two motions of the ball cause different velocities of air to rush past the ball. Now, on the side of the ball turning into the rush of air, the total velocity of the air is slowed down because the spinning air partly counteracts the speed of the air rushing past. On the side of the ball turning away from the rush of air, the velocity of the spinning air and the velocity of air flowing past the ball are in the same direction. Therefore, the velocity of the air is greater on the side turning away from the rushing air.

Keeping in mind that Bernoulli's theorem states that total energy in a system remains constant, it becomes evident that if the velocity of air over a surface is increased, the pressure exerted by the air on the surface must decrease, thus keeping the total energy constant. Conversely, decreased velocity of air over a surface must result in increased pressure on the surface.

Returning to the pitched baseball, you now can see that unbalanced pressure would be exerted on the spinning ball. On the side of the ball turning away from the rushing air, the pressure would be decreased because the velocity is increased, while on the other side the pressure is increased because the counteracting air velocities have decreased the total velocity.

The building up of the lift force on an airfoil differs only in application from the pitching of a curve ball. Notice that the top surface of the wing section shown below has a *greater curvature* than the lower surface.

The difference in curvature of the upper and lower surfaces of the wing builds up the lift force. Air flowing over the top surface of the wing must reach the trailing edge of the wing in the same amount of time as the air flowing under the wing. To do this, air passing over the top surface has to move at a greater velocity than air passing below the wing because of the greater distance the air must travel via the top surface. The increased velocity means a corresponding decrease of pressure on the surface. Thus, a pressure differential is created between the upper and lower surfaces of the wing, forcing the wing upward and giving it lift.



Airflow over a wing section

You now can see that lift force is created by a change of momentum in the air mass. In this discussion lift will be approached from either the standpoint of increased force on the under surface of an airfoil or from the standpoint of increased velocity of the airstream passing over the top surface.

Always keep in mind that whenever an object changes its direction or rate of motion unequal forces must act on it.

Boundary Layer. Boundary layer refers to a condition occurring as a result of friction between an airfoil surface and air moving by it. This clinging of air to an airfoil, such as a wing, is a serious problem in aircraft design. It has been shown that lift depends upon circulation around an airfoil. When circulation is restricted, lift is reduced. Experiments along such lines as highly polished surfaces are constantly reducing these unwanted surface disturbances.

Now that you are familiar with the basic laws of physics which are involved in flight, you are ready to consider other factors contributing to aerodynamics.

PROBLEMS OF AERODYNAMIC FORCES

A discussion of the problems of aerodynamic forces involves several flight terms which you need to understand. So before we take up the specific problems, a few of these terms will be explained.

Basic Flight Terminology

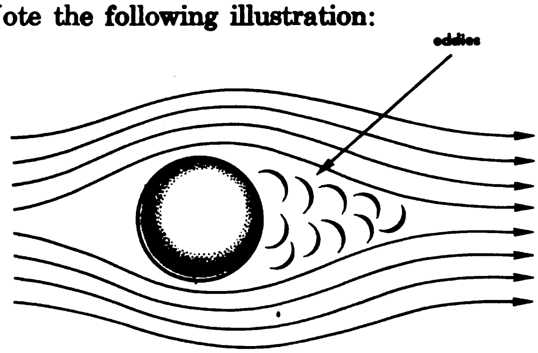
The following explanations of flight terms are intended to give the basic meanings of the terms; they are not an engineer's definitions.

AIRFOIL. An airfoil may be described as any structure around which air flows in a

manner that is useful in flight. The most important of the airfoils on an aircraft are the wings. Other airfoils of an aircraft consist of tail surfaces and the fuselage.

DRAG. The resistance of an object to the flow of air around it is called drag. Drag is due in part to the adhering of air to a surface (boundary layer) and in part to the piling up of air in front of the object. One of the main objectives in aerodynamics is to reduce this resistance yet maintain a high amount of lift and stability.

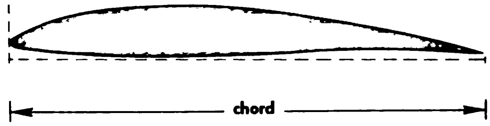
STREAMLINES. The paths of air particles as they flow past an object are called streamlines. Note the following illustration:



Streamline flowing past a spherical body

WING SPAN. Wing span is the measured distance from the tip of one wing to the tip of the other.

CHORD. The distance from front to back of a wing is called the chord. See the illustration below. The chord usually varies in length along the wing. Therefore, it is usually necessary to speak of the *average* chord. The average chord is considered to be the ratio of the wing area to the wing span. For example, if a wing has an area of 48 square feet and a span of 16 feet, the average chord would be determined by dividing the area (48 sq ft) by the span (16 ft). The average chord would be 3 feet.



Measurement of chord on a wing section

CAMBER. Camber refers to the rise of a curve of an airfoil section. Camber is usually expressed as the ratio of the departure of the curve from a straight line joining the extremities of the curve to the length of this straight line. Upper camber refers to the upper surface, lower camber to the lower surface, and mean camber to the mean line of the section. Camber is positive when the departure is outward and negative when it is inward. Note the figures on the following page.

ATTITUDE. Attitude refers to an airborne craft's position in relation to the ground.

STABILITY. A *stable* body is one which returns to its initial position after it has been disturbed by some outside force. If outside forces disturb a stable aircraft from its normal flight, the aircraft tends to return eventually to its original position.

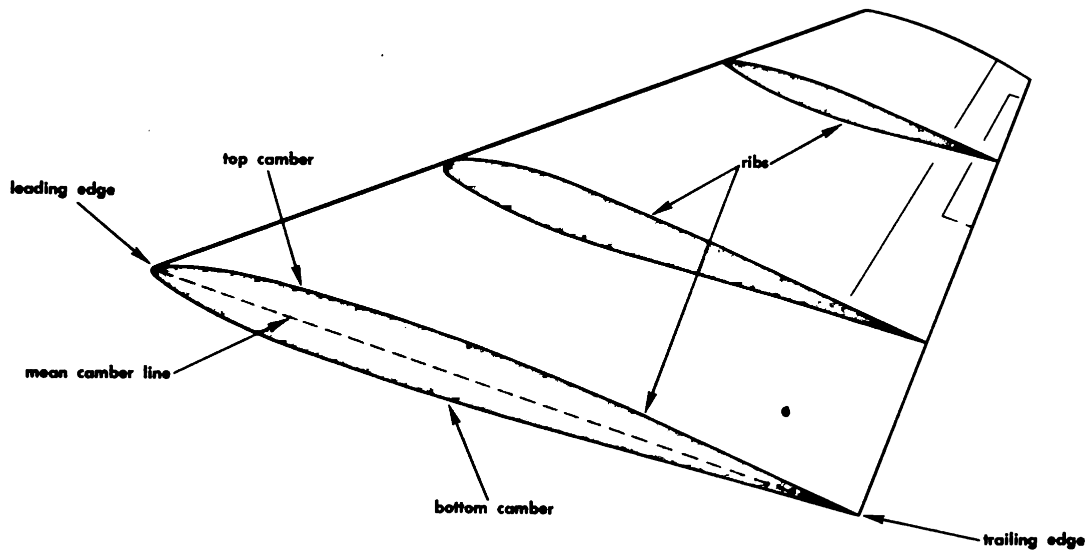
Sometimes a body, if disturbed from its original position, assumes a new position and neither returns to its origin nor moves any farther from it. Such a body is *neutrally stable*. If the attitude of a neutrally stable aircraft is changed by an outside force or change in controls, the aircraft does not tend to return to the original position. It remains, instead, in the new position until other forces influence it.

A third type of stability is negative stability or *instability*. In a case of instability, a body displaced from its original position tends to move farther away. For example, if an unstable aircraft is put into a climb, it tends to climb more and more steeply until it stalls.

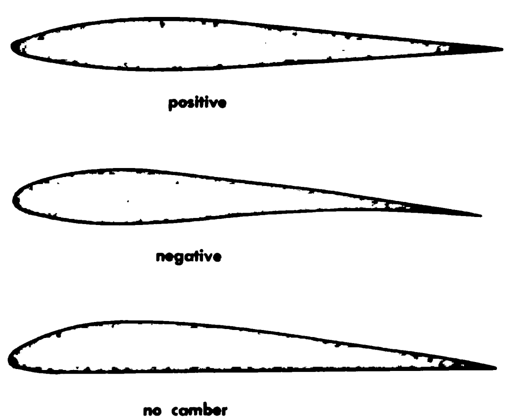
AXIS. A missile moves about three axes as shown in the drawing on page 22. In normal level flight, the vertical line if considered as the *yaw axis*, the longitudinal line through the center of the fuselage is called the *roll axis*, and the line that is perpendicular to the line through the fuselage and parallel to the wings is called the *pitch axis*. Whenever there is any displacement of the missile about any of these axes, the outcome may be any one of the following actions:

1. It may oscillate about the axis and return to its initial position.

2. It may increase its displacement and get out of control.



Camber of air foils



3. It may return to its original position readily without any oscillation.

The last possibility, which indicates a stable body, is desired. We will see later how this problem of stability is met.

DIHEDRAL AND CATHEDRAL. A dihedral angle is the angle formed by a reference line through the wing surface and the lateral axis of the aircraft. This angle lies in a plane perpendicular to the longitudinal axis. Cathedral is the angle that the wings make in a downward direction toward the wingtips from the fuselage. Cathedral is often used on craft designed for supersonic speed. Study the illustration at the bottom of the next page.

Problems Pertaining to the Force of Air

Now that you are familiar with some aerodynamic terminology and have a general idea of the forces and actions that are taking place on an airfoil, let's take up some of the more specific problems of air forces. Since a wing section is the most important of the lifting surfaces, much of the following discussion centers on this section.

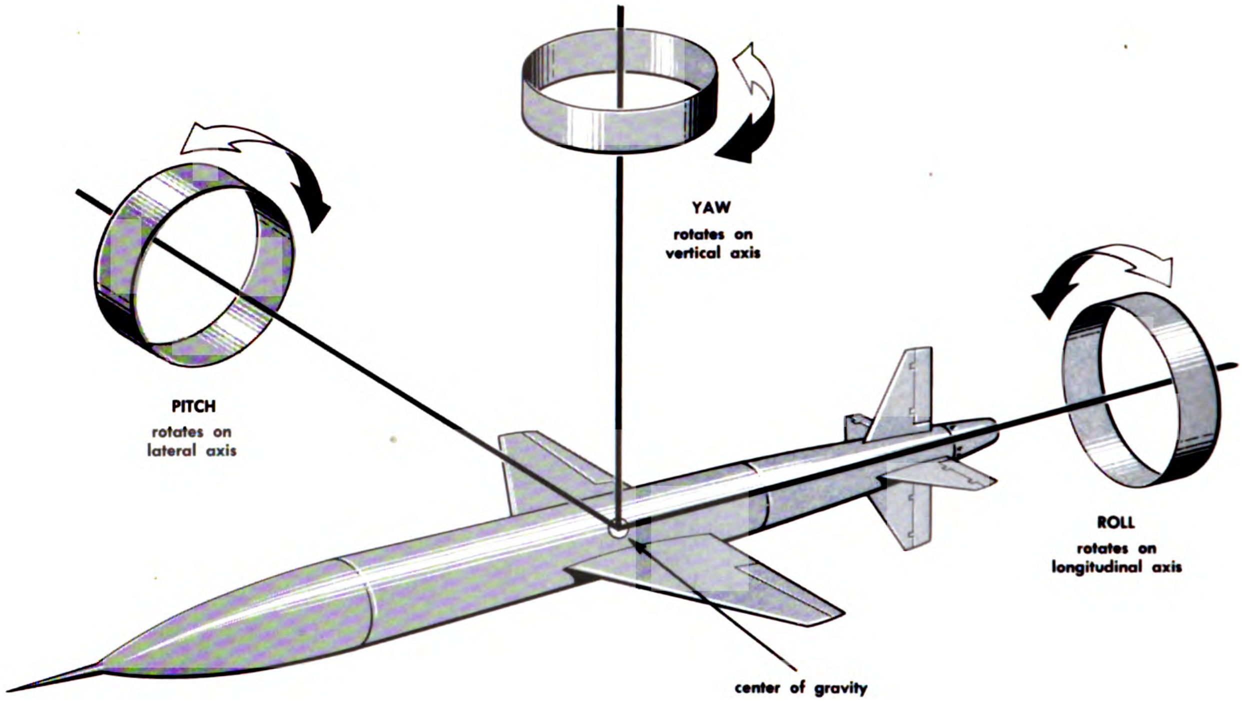
RELATIONSHIP OF MODEL MISSILE TO FULL-SIZE MISSILE. Although a model made to perfect scale flies in a wind tunnel under nearly the same conditions as a full-size missile in the air, the airflows related to each are not the same. This difference is present because the air-mass force and the frictional forces do not follow the same laws. The air-mass force follows what is called the *square law* while the friction force follows the *direct law*. The mass force is proportional to:

$$(\text{Length}) \times (\text{Length}) \times (\text{Velocity}) \times (\text{Velocity}).$$

The friction forces are directly proportional to:

$$(\text{Length}) \times (\text{Velocity}).$$

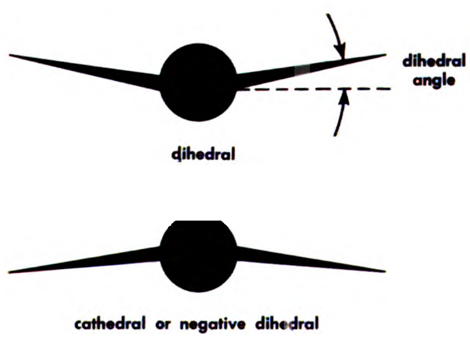
A true relationship hardly ever exists in model tests because the forces on a model generally are much less than the forces on a full-size missile. Because of such discrepancies



Flight attitude of Missiles

between model tests and full-size operation, it is necessary to build full-scale test aircraft to get a true indication from flight test. Differences between the effects of air forces on models and on full-size aircraft become more prominent at supersonic speeds.

CENTER OF PRESSURE. On each infinitesimal part of a wing surface there is a small force present. This force is different in magnitude and direction from the force acting on any other area forward or rearward from this point.



Dihedral and cathedral

It is possible to add mathematically all of these small forces. The sum of all the tiny forces over this surface is called the resultant. This resultant has *magnitude*, *direction*, and *location*. The point of intersection of the line of direction with the chord is called the *center of pressure*.

In actual flight there is a different airspeed for different angles of attack. But if for test purposes the velocity of the airstream is maintained constant while the angle of attack is changed, the results on a nonsymmetrical wing are as shown in the sketches at the right. The sketches show a wing section at various angles of attack and the effect of the different angles of attack on the resultant force and the position of the center of pressure.

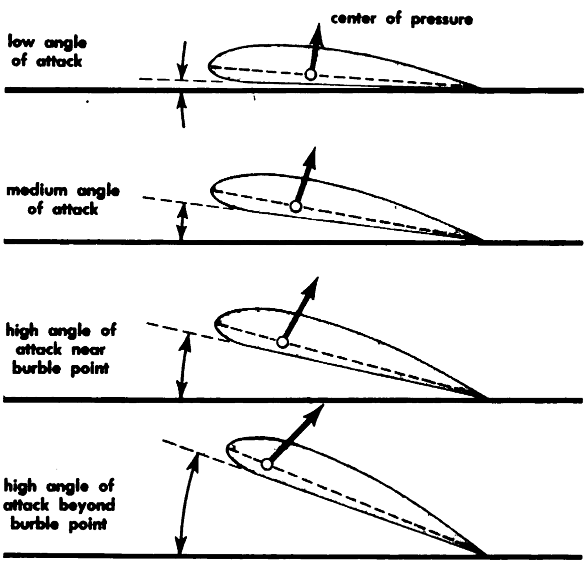
The burble point referred to in the lower sketch at right is a point at which airflow over the upper surface becomes rough, causing an uneven distribution of pressure. The burble point is generally reached when the angle of attack has been increased to 18° or 20°.

At small angles of attack, the resultant is comparatively small. Its direction is upward and back from the vertical, and its center of pressure is well back from the leading edge. You should note that the center of pressure changes with the angle or attack, and the resultant has an upward and backward direction. At a positive angle of attack of 3° or 4°, the resultant has its most nearly vertical direction. Either increasing or decreasing the angle causes the direction of the resultant to be farther from the vertical.

To determine the amount of force that is applied to a surface, it is first necessary to determine the mass of air that is working on that surface. The velocity of the air and the area that it is working on must be known to find the mass of air. Thus, the mass of air is determined as follows:

Mass = (Cross section) × (Velocity) × (Density), where density, by simple definition, is the quantity per unit volume.

Suppose we have a mass of air moving in a horizontal direction that is slightly deflected downward at a certain velocity in feet per second. The lift that the mass of air exerts in pounds is equal to the product of the mass multiplied by the change in velocity:



Changes of center of pressure as a result of varied angles of attack

Lift = (Cross section) × (Velocity) × (Density) × (Deflection velocity). Here we have a simplified formula to compute lift. This formula is basic; you will see later that there are other factors that must be taken into consideration in the computation of lift.

As you learned before, the resultant force on a wing under a particular condition can be described as two components in two directions, the chosen directions being perpendicular and parallel, respectively, to the relative wind. These components, you may recall, were called lift and drag.

DETERMINATION OF LIFT. Lift force depends on the contour of a wing, angle of attack, air density, area of the wing, and the square of the airspeed. The common equation for lift is given as:

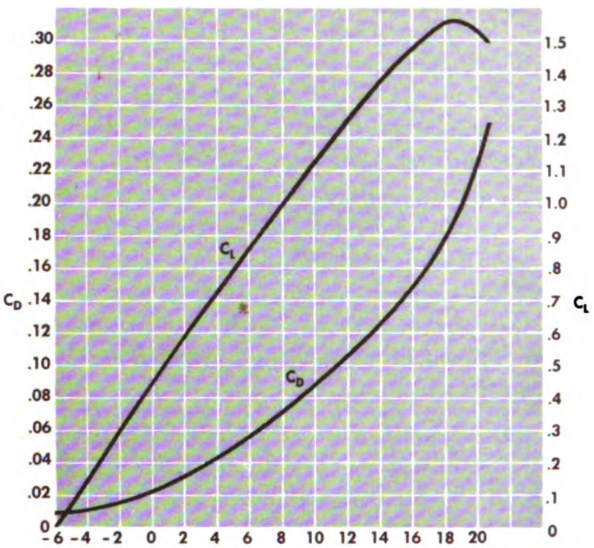
$$L = C_L \frac{\rho}{2} SV^2,$$

where "L" is the lift in pounds, "C_L" is the lift coefficient which depends on the wing contour and the angle of attack, ρ is the air density in slugs per cubic foot, "S" is the area of the wing in square feet, and "V" is the velocity (airspeed) in feet per second. (A slug is defined as the mass which would receive an acceleration of 1 foot per second per second when it is acted upon by an unbalanced force of 1 pound.)

The coefficient of lift (C_L) and the coefficient of drag (C_D) are determined by wind-tunnel tests and are plotted as a characteristic curve for the particular airfoil under consideration. The chart on the following page shows a typical set of curves on a given airfoil.

In order to have level flight, total lift must equal the weight it is supporting. As the angle of attack increases, the lift coefficient increases to a certain maximum value. The maximum value of lift coefficient is the point where the air no longer flows evenly over the wing surface but tends to break away. This breaking away (the burble point) is called the stalling angle. After the stalling angle is reached, the lifting force is rapidly lost, as is the airspeed.

The least possible airspeed occurs when flying at the angle of the maximum-lift coefficient. Another name for this angle of maximum-lift coefficient is, therefore, the



Characteristic lift (CL) and drag (C_D) curves of an airfoil

angle of minimum speed. When weight is added to an aircraft, wing loading is increased. Consequently, the airspeed must be increased in order to fly at the same angle of attack as could be flown before the weight was added.

"Clipping the wings" is sometimes done to increase the speed of an aircraft. This decrease in wing area means an increased wing loading, and a higher velocity is required for any angle of attack as compared with the velocity before the wing was reduced.

DETERMINATION OF DRAG. Drag is the resistance of air to forward motion. The *drag component* of the resultant force on a wing is the component *parallel* to the direction of motion. It is this force that resists the forward motion. In horizontal flight, it is the force that must be overcome by the thrust (the force that is pushing the craft forward). The drag force (in pounds) is the product of a drag coefficient, obtained from characteristic curves of airfoils, times half the air density times wing area in square feet times the square of the velocity. The drag force in formula is:

$$D = C_D \frac{\rho}{2} S V^2,$$

where "C_D" is the coefficient of drag, ρ is air density, "S" is the area, and "V" is the velocity (in feet per second). For small angles of attack, the drag coefficient changes very little with angle of attack. As the angle of attack increases, the drag coefficient increases. The drag coefficient is usually quite small in comparison with the lift coefficient.

In addition to the wings on a missile, the fuselage, tail airfoils, and other surfaces resist motion. *The resistance of the parts of a missile which do not contribute to lift is called parasite drag.* That part of the drag of the airfoils which contributes to lift is called induced drag. Both parasite and induced drag vary as the square of the velocity. The function of the driving force is to furnish a forward-acting force of thrust to balance the backward-acting force of drag.

A force producing an acceleration is an unbalanced force; it is not counteracted by an equal force acting in the opposite direction. Such a force acting on a body produces an acceleration in the direction of the force. If a body is in motion at a certain speed and the forward-acting force of thrust is just equal to the backward-acting force of the total drag, then there is no acceleration; the velocity, in other words, remains the same.

If the thrust is decreased in magnitude so that drag is greater than thrust, an unbalanced backward force results. This causes a backward acceleration, or deceleration, of forward velocity. Decreasing the forward velocity causes a decrease in drag. When the drag force has decreased in magnitude until it is equal to thrust, there is no further deceleration and the body moves forward at the new, slower velocity.

Wing drag varies with air density and as the square of the velocity of the missile. In level flight, a missile flies faster at altitude than at the same angle of attack at sea level. This occurs because air density is less at higher altitudes; thus the wing travels faster because of the decrease in density. The effect of decrease in density on the wing is exactly neutralized by the increase in velocity squared. Therefore, for a given aircraft, the wing drag depends on the angle of attack and is independent of altitude. The known drag for one altitude is essentially the same at any other altitude, provided the angle of attack remains unchanged.

To maintain the forward movement of the wing through the air, a force equal to the drag must be constantly exerted. This force multiplied by the velocity is equal to the power which must be expended in maintaining forward motion. If the force in pounds is multiplied by the velocity in feet per second, the product is power in foot pounds per second. By definition, *horsepower* is 550 foot pounds per second. The horsepower needed to move a wing through the air is therefore:

$$hp = \frac{DV}{550}$$

where "D" is the drag in pounds and "V" is the velocity in feet per second.

Aspect ratio. In discussing the forces on a wing, the area of the wing has been mentioned frequently. An area may be in the form of a square or a rectangle with different ratios of length of sides. There is a difference in the forces on a wing depending on the shape of the wing, even though the area remains the same.

We can set up a relationship between the *span* and the *chord* of the wing. This relationship is called the aspect ratio. In a rectangular-shaped wing, it is the *ratio of span to chord.* If the wing is not a simple rectangle but is *tapered* or has *elliptic* tips, the span is still the extreme distance from tip to tip but the chord varies from one position to another along the span. It is possible to find an average chord, but it is simpler in the case of a nonrectangular wing to define the aspect ratio as the ratio of span squared to the area. In formula this ratio reads:

$$\text{aspect ratio} = \frac{b^2}{S}$$

where: "b" is the span of the wing
"S" is the area

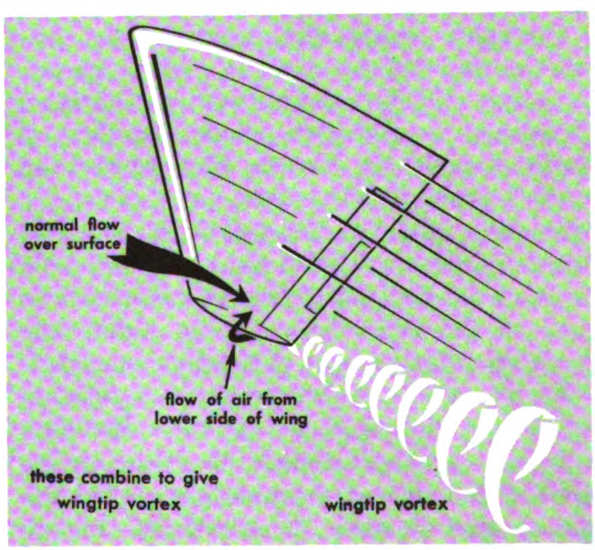
As mentioned, the aspect ratio of a wing must be known in order to fully determine the extent of the forces acting on the wing.

Wingtip vortex. As air flows about a wing, the pressure of the air immediately above the upper surface is less than the air pressure immediately below the surface. With the air at higher pressure below the wing, air will spill by the wingtips to the upper surface as shown in the figure to the right.

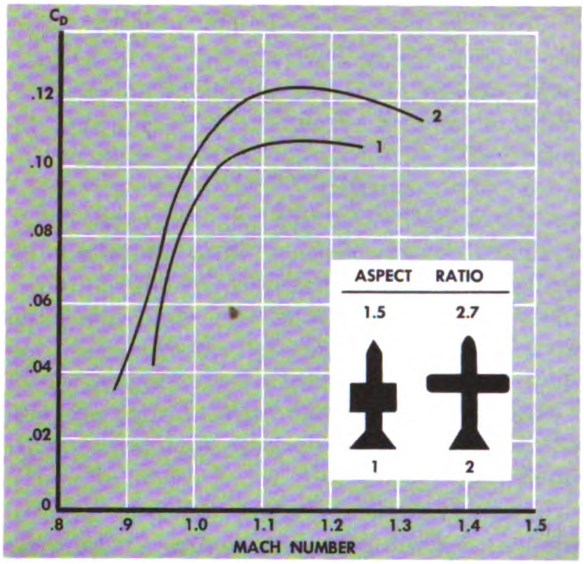
This flow of air from the lower surface combines with the normal flow of air, causing a swirl of air at the wingtips. The swirl is called a *wingtip vortex.* At each wingtip the action of the vortex is to throw the air inward and downward. The *downflow* caused by the wingtip vortices has a greater effect in disturbing the ordinary flow about a wing with a shorter span than a wing with a longer span. For two wings of the same area, the wing with the smaller aspect ratio has a shorter span than a wing with a larger aspect ratio. Thus, wings with large aspect ratios are less affected by wingtip vortices than wings with small aspect ratios.

In analyzing the effect of the wingtip vortices, we will consider the drag of the wing as being made up of two parts, called profile drag and induced drag. Profile drag is essentially the skin friction of air on the wing and is practically independent of the angle of attack and the aspect ratio. Induced drag is related to the downflow of the wingtip vortices. The magnitude of induced drag depends on both the aspect ratio and the angle of attack.

Since the *total* drag of a wing is the sum of its *profile* drag *and its induced* drag and since the induced drag changes with various aspect ratios, the total drag changes with various aspect ratios. With higher aspect ratios, the effect of the wingtip vortices is to give the



How wingtip vortices are formed



Effect of aspect ratio on wing drag

same lift coefficient as smaller angles of attack. The net result of changes in aspect ratio are shown in the above illustration. For a given aspect ratio at any one angle of attack, the lift coefficient is greater and the drag coefficient is less than for a wing of smaller aspect ratio.

The preceding figure shows that from a theoretical standpoint the biggest possible aspect ratio is most desirable, but for practical reasons of structural strength the aspect ratio is seldom greater than 8.5 for aircraft.

CONTROL ABOUT THE THREE AXES. Aircraft must be constructed in such a manner that they will fly a course without continual correction in direction. Such stability is made possible by devices that control an aircraft about its three axes.

Stability about the Vertical Axis. Stability about the vertical axis is commonly provided for by a vertical fin. If an aircraft tends to turn to the right (yaw to the right), the pressure on the left side of the fin is increased. This increased pressure resists the rotation and forces the tail in the opposite direction.

In conventional aircraft, the fin may be divided and have a movable part called the rudder that is used for directional control. In addition to the rudder, there may be trim tabs that can be set for a particular direction of flight relative to prevailing air currents. The tabs compensate for the air forces, doing away with the need for continually changing the rudder position.

Along with the fin, the vertical sides of the fuselage act as stabilizing surfaces. The same action takes place here as on the fin.

Another means of obtaining yaw stability is by sweepback of wings. *Sweepback* is the angle that the leading edge of a wing makes with the longitudinal axis of the missile measured from the nose. If a missile yaws to the right, the leading edge of the left sweepback wing becomes more perpendicular to the relative wind while the right wing becomes less perpendicular to the relative wind. This condition places more drag on the left wing and less on the other. The unbalanced drag tends to equalize, thus forcing the missile back to its original attitude.

Stability about the Longitudinal Axis. Stability about the longitudinal axis is provided for in a missile by dihedral and the position of the wing.

Dihedral produces stability by causing a change of lift on the wing surfaces. As a missile starts to roll, it will sideslip slightly and thus create a relative wind component. This component increases the lift on the lower wing and decreases the lift on the higher wing. Lift increases on the lower wing because the angle of attack becomes greater on the wing.

The positioning of the wings at the time an aircraft is constructed is another means of obtaining stability about the longitudinal axis. An aircraft has greater stability if the wings are placed above the center of gravity than if they are placed below the center of gravity. Aircraft generally are classified according to four wing types: low-wing, mid-wing, high-wing, and parasol-wing.

Stability about the Lateral Axis. Stability around the *lateral* axis is called *pitch stability.* Pitch stabilizing is accomplished by a horizontal surface at the tail of the aircraft. This surface is known as the stabilizer. The stabilizer may be considered as consisting of two sections: the stationary part as the stabilizer and the movable part as the elevator. Stability with reference to pitch is

accomplished by the increasing forces present on the tail surface when the aircraft changes its angle of attack. For example, if a missile tends to nose downward, the force on the upper surface of the stabilizer would increase, forcing the tail down and bringing the missile back to its original attitude.

Elevation tabs are used to correct for any existing factors causing *pitch*, just as the trim tabs are used in the case of the rudder.

Other types of control surfaces and stabilizing devices are taken up later in this chapter.

AERODYNAMICS OF SUPERSONIC FLIGHT

You should now have a basic understanding of the forces that are constantly working on an airframe. Such information will help you to understand the new problems that arise in supersonic flight. The major problems in high-speed aerodynamics are listed as follows:

a. Drag reduction.

b. Lift effectiveness.

c. Aerodynamic control.

d. Stability: static, dynamic, automatic.

e. Propulsion aerodynamics.

f. Aerodynamic loads.

g. Atmospheric conditions.

h. Techniques for studying high-speed aerodynamics.

Up to the end of World War II, improvements in aircraft performance and speed came slowly. Progress was made primarily in the development of larger engines and minor aerodynamic changes in former design. As jet power plants came into prominence, the speeds of new aircraft could be increased greatly. The old design procedures, formerly satisfactory, were changed for several reasons. For one thing, the flow of air at high speeds follows a different set of laws. Near-sonic speeds meant dealing with compressibility and Mach-number effects. At these speeds there is the problem of heating to be dealt with. It also meant changes in wind-tunnel design, scale effects, types of airfoil sections, problems of stability, and other factors that up to this time had little significance.

The next pages introduce you to some of these problems and try to show how the problems were compensated for by changes of design. As before, it is first necessary to become familiar with new terminology.

Supersonic Flight Terminology

Several terms which you need to understand for a discussion of supersonic flight are explained in the following paragraphs.

SHOCK WAVE. As the speed of a missile increases, there comes a point at which the air can no longer get out of the way fast enough. The air tends to pile up or compress in front of the missile. This piling up forms pressure impulses which can travel at the speed of sound because of the compression of air. However, when the missile reaches the sonic speed, the air can no longer separate before it. At this point the air is split sharply, setting up what is known as *shock waves*.

In a shock wave the pressure of air varies sharply, seriously altering the forces and pressure distribution on a missile. Strong changes in trim are necessary. The flow of air over the wings separates, much as in a low-speed stall. Tail surfaces may get a good buffeting. Wing drag rises. And deflection of control surfaces may cause new shock waves, which interact so that the controls may become ineffective at certain speeds.

MACH NUMBER. Since the speed of sound plays a big role in determining shock waves and air-flow characteristics at high speeds, the ratio of flight speed to the speed of sound is significant. This ratio is called Mach number in honor of an Austrian scientist, Ernst Mach (pronounced "mock"), who first pointed out its importance in 1887. If a missile travels at a speed twice the speed of sound, the missile has a flight speed of Mach 2.0. If a missile has a speed half that of sound, it has a flight speed of Mach 0.5. The speed of sound varies with altitude, decreasing from 760 miles per hour at sea level to about 675 miles per hour at 30,000 feet.

REYNOLDS NUMBER. In dealing with scale models and actual size aircraft, the results of tests may not be comparative even after scaling down all known variables. In some cases the variables have entirely different or opposite effect when related to the scale model. This difference is true especially in the case of finding the coefficients of lift and drag.

The coefficients of lift and drag vary greatly depending upon the ratio of values of the model size, relative wind speed, and air viscosity and density. These variables have been put into a mathematical proportion that is called *Reynolds* number. Only when the value of the ratios of these variables are the same are the tests of the model and the actual aircraft comparative.

HEAT BARRIER. As air is compressed by the ram effect at high speeds, a temperature rise takes place. The energy of the moving body transformed into a temperature is known as the *heat barrier*. The temperature rise is *directly* proportional to the square of the supersonic velocity. Therefore, as the speed of sound is doubled the ram temperature is increased four times. The standard temperature at sea level is considered to be 59° Fahrenheit. This temperature decreases with altitude up to 46,500 feet, after which it is assumed to be constant. At sea level ram temperature is about 88° F at 760 miles per hour, 29° hotter than the standard sea-level temperature. Ram temperature is about 260° F at 1300 miles per hour, and about 1000° F at 2600 miles per hour. As flying speeds increase, new materials to withstand the high temperatures will be needed.

SPEED CLASSIFICATIONS. So far, mention has been made of two general terms when relating speeds to sound: *subsonic* and *supersonic*. However, the speed categories can be classified more specifically as subsonic, transonic, supersonic, and hypersonic. Speeds are called subsonic when a body moves through air at a speed less than the speed at which the pressure wave is building up. That is, air in front of the body has a forewarning of the approaching body and is able to get out of the way.

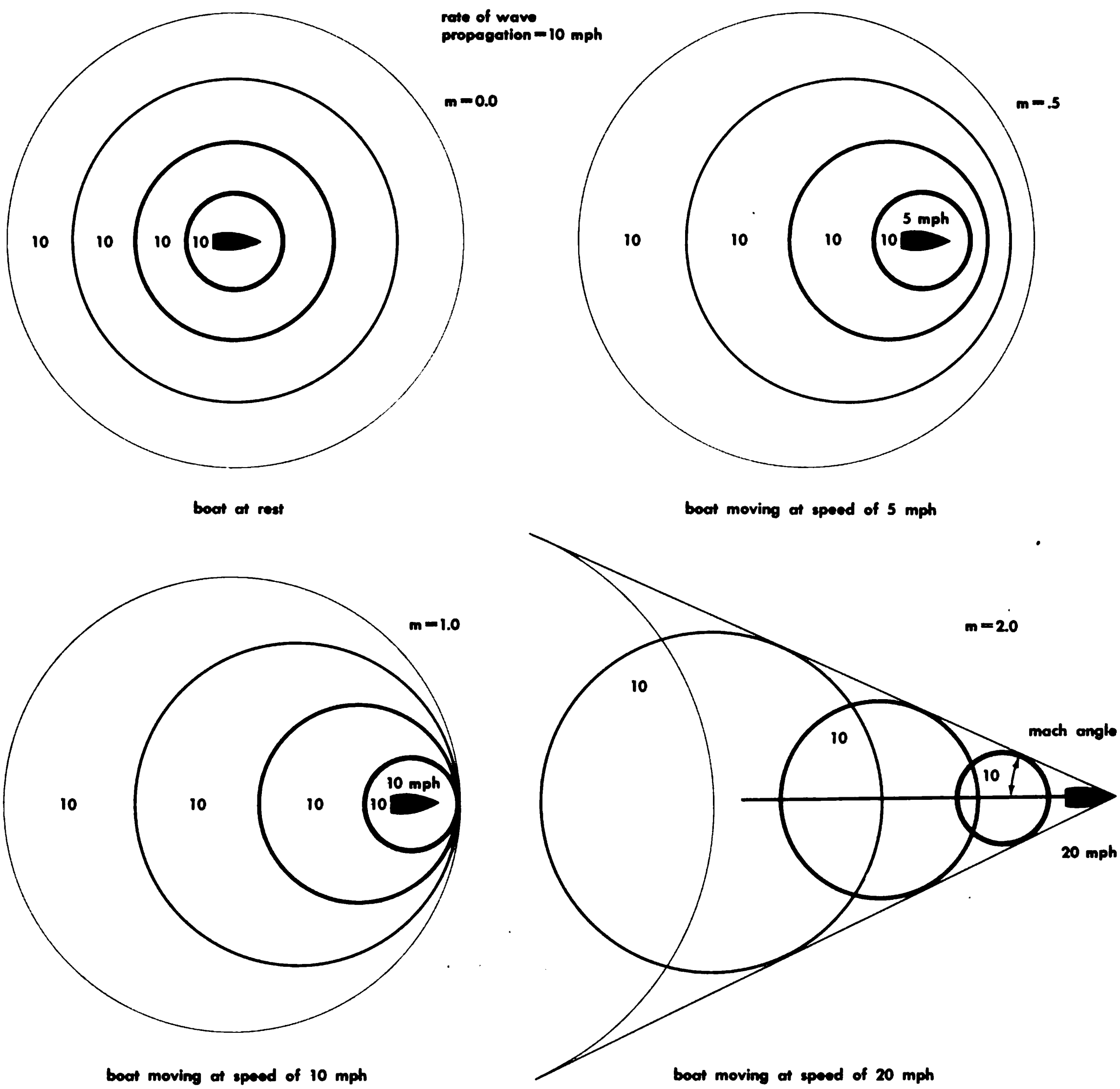
When a point is reached at which the speed of the body is the same as that of the pressure wave, waves build up into a single pulse of greater magnitude. This condition exists when the body moves at the speed of sound; the body is then at sonic speed. It is at this speed where transonic speeds also occur; that is, when a body is moving at sonic speed, there are points at which air velocities about the body are above the speed of sound and there are points at which air velocities are below the speed of sound. Considerable turbulence and buffeting of missiles takes place in this region; therefore, it is desirable to pass through the transonic speeds rapidly to prevent the unwanted disturbances.

Supersonic speed is present when the air flow at all points about the body is greater than the speed of sound. In supersonic flow, little turbulence is present.

As mentioned, the fourth classification of speed is hypersonic, sometimes called ultrasonic flow. As a body moves through the air at high speeds, a short amount of time is necessary for the molecules of air to adjust themselves to the presence of the body and to readjust themselves after the body has passed. This period of adjustment and readjustment is termed the *relaxation time*. If a body is moving at a speed greater than the relaxation time, it is in a new velocity range which is called *hypersonic*.

MACH ANGLE. To explain what is meant by Mach angle, picture a boat on a lake. If the boat is at rest and is rocked by a wind, ripples will radiate from the boat in all directions. The speed at which these ripples travel is called the propagation rate. Assume that these ripples were propagated at a speed of 10 mph. The ripples would be concentric, and each would remain inside the previous one.

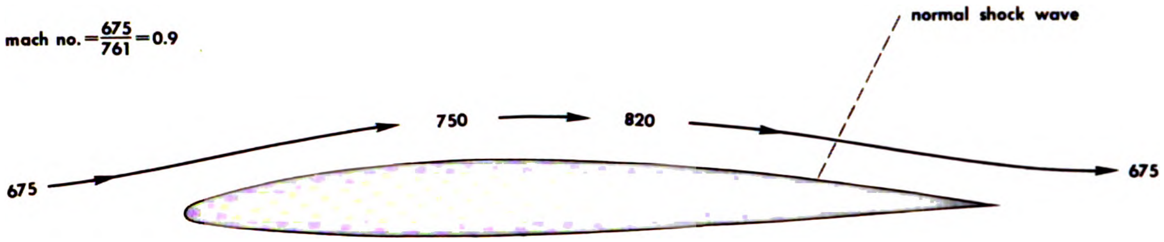
Now if the boat moves forward with a speed of 5 mph, the ripples will still radiate at 10 mph. Each circle will stay inside of the previous one, but they will not be concentric. The ratio between the boat speed and the propagated wave speed is 5/10. If the propagation rate of the ripples from the boat is thought of as being the speed of sound, the above relationship of 5/10 corresponds to Mach 0.5. If the boat were to increase its speed to 10 mph, then the relationship corresponding to the speed of sound would be Mach 1.0. Now assume that the speed of the boat is increased to 20 mph. This relationship corresponds to Mach 2.0. The ripples still form at 10 mph, but the center of disturbance is moving twice as fast. The wave pattern now becomes a wedge on the surface of the water. In the air, with three dimensional flow, the pattern would be a cone. The semivertex angle is the *Mach angle*. This angle is illustrated on the following page.



Analogy of Mach numbers and Mach angle

NORMAL SHOCK WAVE. Pictured at the top of the following page is a normal shock wave. The airfoil section has a speed (the free-stream velocity) of 675 miles per hour. The figure indicates the speed of sound to be 761 miles per hour. Dividing 675/761 gives a Mach number of 0.9. As the air gains speed over the airfoil, it increases to 750 mph and then to 820 mph, which is supersonic. The air then begins to slow down to its original speed of 675 mph. Now if you will recall, Bernoulli's theorem says that if the velocity of air changes, its pressure changes, Since the velocity has changed rather suddenly from 820 mph to 675 mph, we could expect a large pressure difference to exist. The boundary between the two areas of pressure difference is where the *normal shock wave* occurs. Normal shock waves absorb a large amount of energy which is dissipated in the form of heat.

OBLIQUE SHOCK WAVE. In supersonic flight there occurs an additional type of wave called an oblique shock wave. It has this name because it meets the free-stream at an oblique angle, an angle greater than 90°. An oblique shock wave is shown on the following page. The component of flow perpendicular to the wave must go from supersonic to subsonic



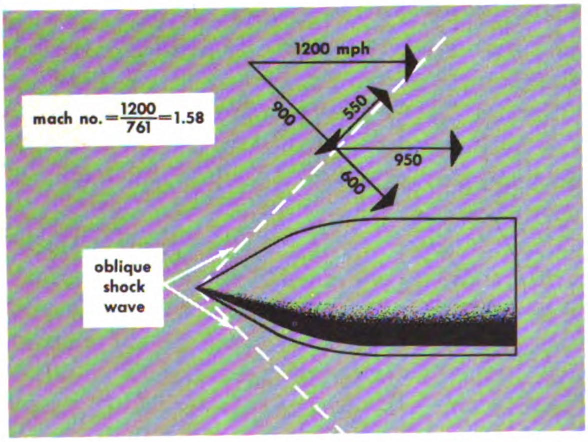
Normal shock wave

velocity just as in the case of the normal shock wave. If we assume these values to be 900 mph to 600 mph, the free-stream velocity may be 1200 mph and the component parallel to the shock wave may be 550 mph. The resultant velocity (say 950 mph) behind the wave is equal to the vector resultant of the 1200- and 550-mph components. This resultant velocity always flows parallel to the surface of the body, and it may or may not be supersonic. The original velocity of the free-stream and the oblique angle of the shock wave are the two factors that determine the final velocity.

Now that you have been introduced to the meaning of several terms that pertain to high-speed flight, you are ready to take up the problems of supersonic flight.

Guided Missile Configuration

The configuration of a missile is a primary concern of missile designers, because a missile's configuration largely determines the extent of drag and lift acting on the missile. And these two forces in turn largely determine the overall efficiency of the missile.



Oblique shock wave

Drag reduction. The achievement of low drag in supersonic flight is of great importance. With low drag configuration, smaller missile power plants can be used and less fuel capacity is necessary.

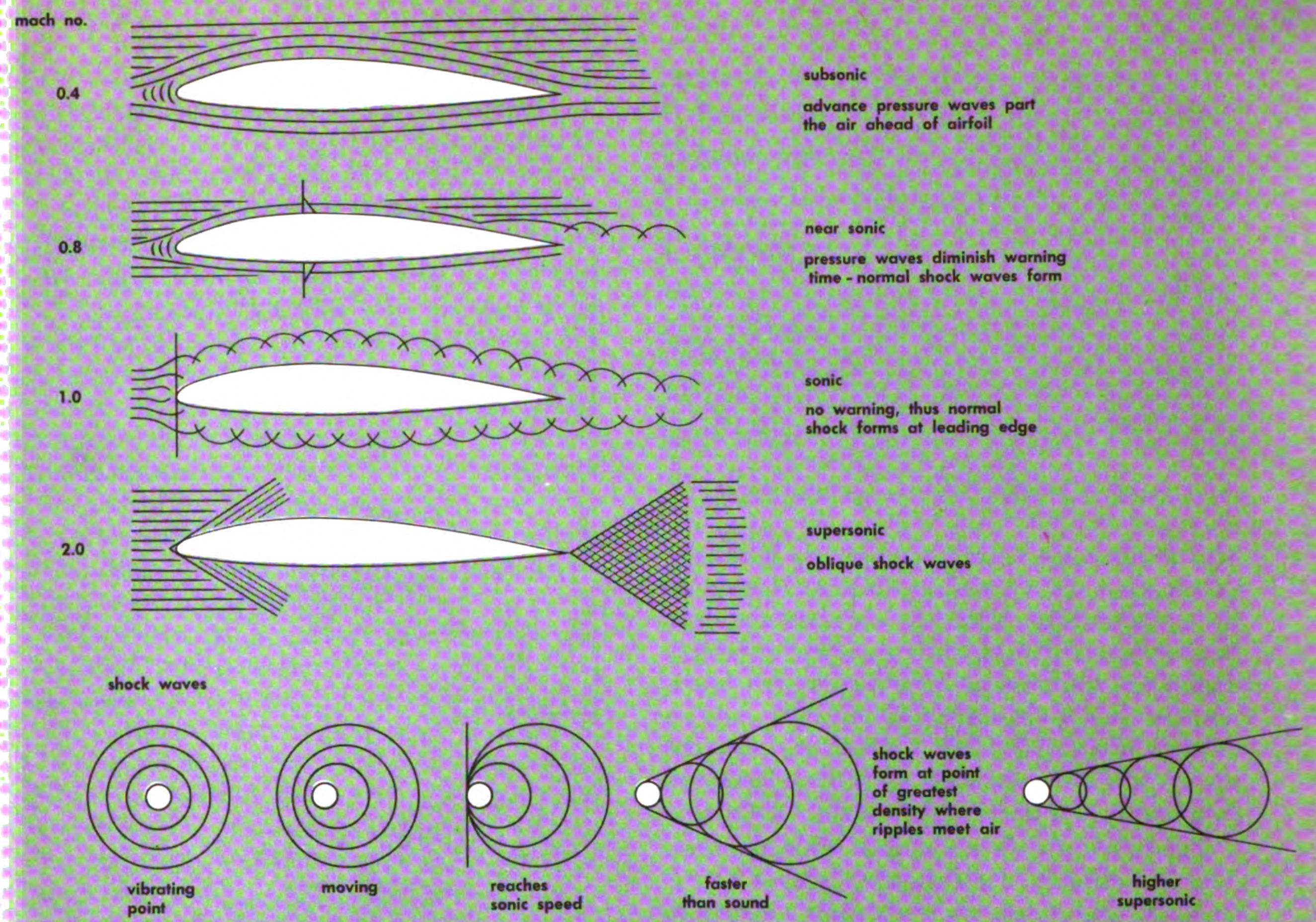
The drag of a missile is made up of fuselage drag, wing and fin drag, and drag developed by the interference of the various drags. The effects of thickness distribution, Reynolds number, surface imperfection, and Mach number all influence the drag. Wing drags also are greatly influenced by thickness ratio, sweepback, aspect ratio, and section of airfoil.

Contrary to subsonic experience, the total drag of a supersonic missile is not necessarily the sum of the separately measured drags of its elements. For example, the drag of a wing seems to be determined by the body shape on which it is mounted.

A simple illustration of an interference effect in supersonic air flow is given in the graph on page 32. The curves were obtained with free-flight tests of rocket-powered models.

The drag of a blunt-nosed body is shown by the upper curve. Adding the conical point, which in itself has a positive drag, greatly reduces the drag of the combination of conical point and body. Apart from overcoming interference effects, windshilds of the type shown may be useful in cases where optimum aerodynamic shapes must be compromised, as would be the case when forward vision of a seeking unit requires a hemispherical nose shape to avoid distortion of signals.

Lift effectiveness. A *lift force* is required for maneuvering. For long-range winged missile, the lift must provide the required support at minimum drag. Lift must also vary smoothly with angle of attack if control response is to



Effects of speed on air

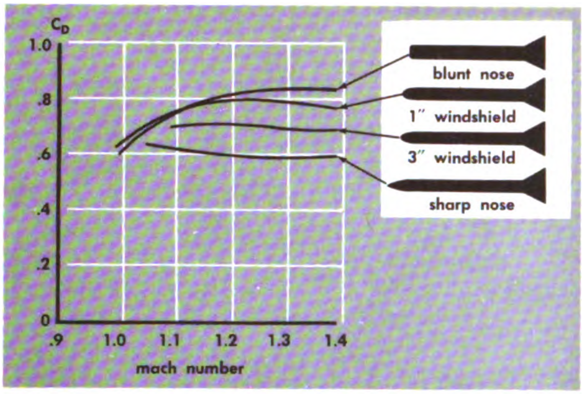
be satisfactory. The lift behavior of three different wing designs is shown in the graph at the bottom of the first column on the next page.

The upper curve (a) shows the behavior of a high aspect ratio, 10%-thick wing with no sweepback. In this case, an abrupt reduction of lift effectiveness occurs slightly above Mach 0.8. However, definite positive lift effectiveness is retained through the speed of sound, into the supersonic range. The second curve (b) shows typical behavior of a relatively thick wing at high speeds. In this case, a complete loss of lift effectiveness occurs at high subsonic Mach numbers. The third curve (c) shows the effect of sweepback in delaying and minimizing the critical changes produced in the transonic range. The sweptback wing of (c) has the same section as the high aspect ratio wing of (a). With the sweptback wing, there is very little variation in the lift behavior over the speed range tested.

AIR FLOW AT DIFFERENT SPEEDS. The conditions of flight associated with subsonic flow are well known. Supersonic conditions appear to be orderly and, consequently, of such a nature as to be readily analyzed mathematically. But in the transonic speed range, major design problems arise.

The drawings to the right on the next page show the air-flow pattern, along with lift and drag changes, as a circular arc wing section moves from subsonic to supersonic speed.

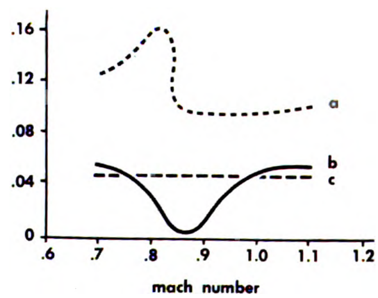
The top figure shows an airfoil in high subsonic flow. The free-stream Mach number is 0.8. This condition is well known and is the kind encountered every day. The lift is much greater than the drag. The second figure indicates a condition of early transonic flow.



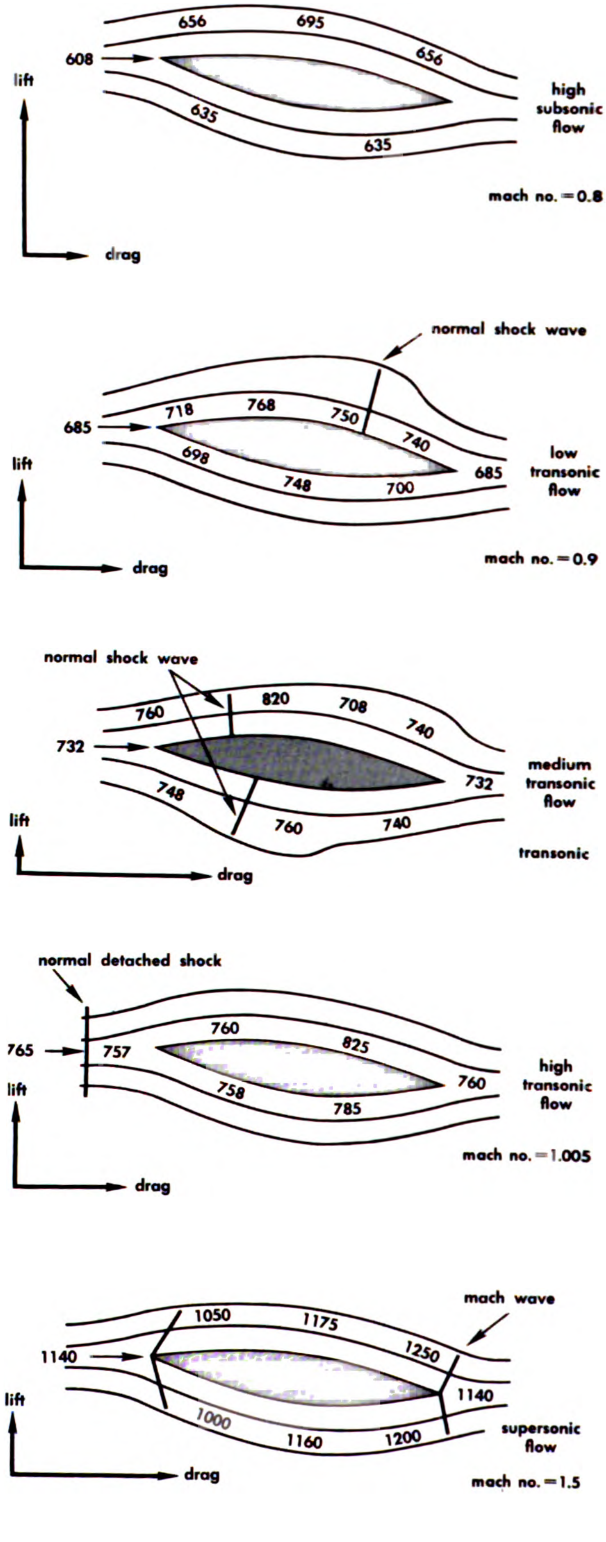
Effects of conical windshields on body drag

The free-stream Mach number has increased to 0.9 and a normal shock wave has formed on the upper surface. Lift has decreased and drag has increased. The third figure is in the middle of the transonic range. The free-stream Mach number has now increased to 0.95, and a normal shock wave has formed on both the upper and lower surfaces. Lift has decreased even more, and the drag has increased sharply.

The fourth figure shows late transonic flow. The free-stream Mach number is over 1.0. A weak normal shock wave has just started to form a few inches ahead of the leading edge. The normal shock waves on the rear upper and lower surfaces have virtually disappeared, since the flow is now sonic over that portion of the section. Lift is about the same, and the drag has decreased a small amount. The bottom figure shows fully developed supersonic flight conditions. The free-stream Mach number is now 1.5. The normal shock wave has moved back and become a Mach wave, and the usual Mach wave has formed on the trailing edge. The air-flow condition is now stable. The



Lift effectiveness



Air flow patterns at different speeds

lift is equal to the weight of the missile. The drag is high but this is caused by the high speed and not by energy losses in the shock waves as was the case in transonic flow.

Although much has been accomplished in regard to the overall picture of lift effectiveness, there still remains much to be learned.

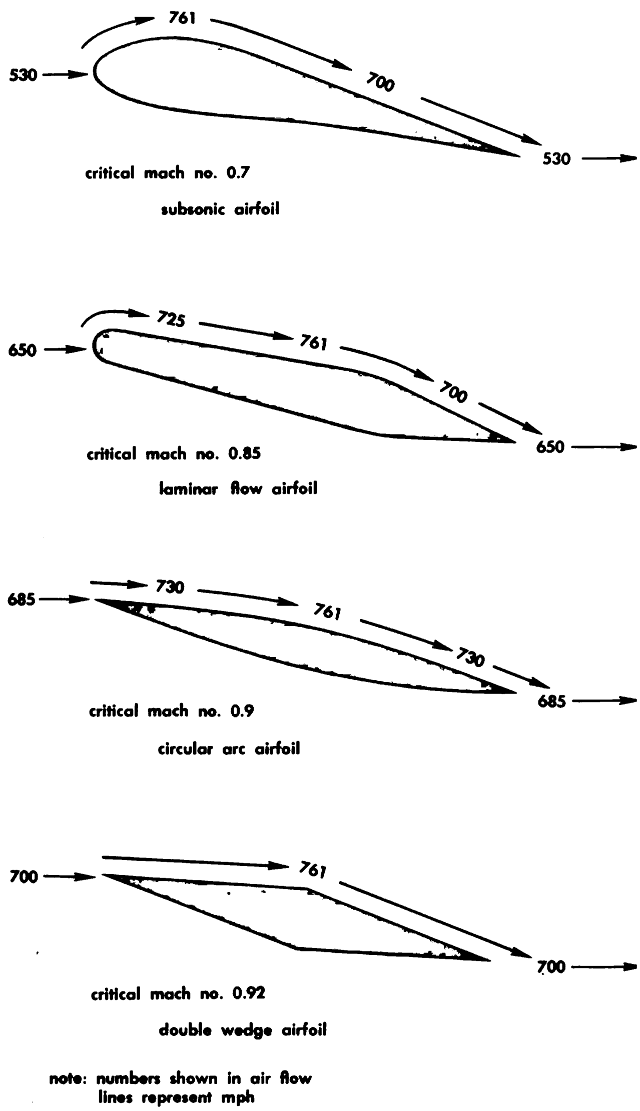
WING DESIGN. Air flow over an ideal wing would be subsonic until a velocity of Mach 1 is reached, and then it immediately would become supersonic. In other words, the transonic range would be eliminated. Actually the *transonic region* begins when the flow over any part of the wing first becomes supersonic. The free-stream Mach number at which transonic flow begins is called the critical Mach number for the wing.

The drawings at the right illustrate the evolution of a high-speed airfoil section. Pictured first is a typical low-subsonic airfoil with a critical Mach number of 0.7 at sea level. Sonic speed occurs first near the leading edge. When the normal shock wave forms, it decreases the lift over most of the upper surface. The center of the lift shifts forward, and drag increases rapidly.

The next sketch is a typical high subsonic airfoil with a critical Mach 0.85. The airflow does not reach its maximum value until it reaches a point far back on the wing. When the normal shock wave does form, the loss in lift and the increase in drag are much less than with the first airfoil.

The third type is a low-supersonic-type airfoil. The critical Mach number may be as high as 0.9. This is known as a circular-arc airfoil. The acceleration of air over the surface is gradual. Because of its symmetrical shape, the center of pressure on the airfoil changes very little when the lift decreases. This characteristic greatly simplifies the control problem.

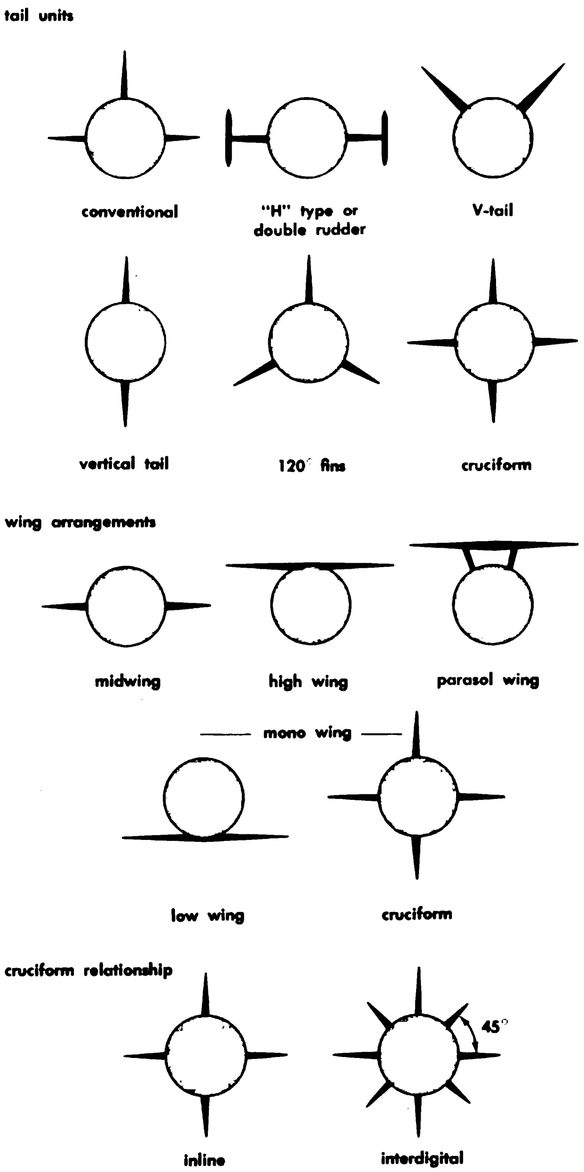
The fourth type is known as a wedge airfoil. It also is a supersonic section having a critical Mach number of 0.9 or better. It can be shown mathematically that the ideal supersonic airfoil is a flat plate; however, such an airfoil would not have sufficient strength to support the load. This wedge then is a *modified flat plate* designed to give more strength than a flat plate.



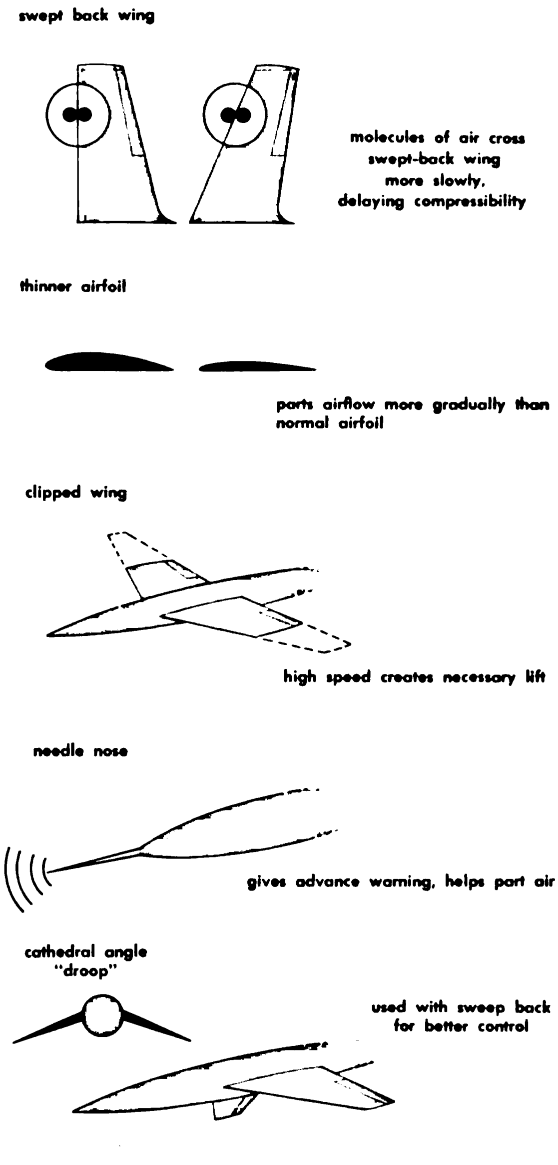
Evolution of high-speed airfoil sections

The low-speed characteristics of the wedge and circular arc airfoil are poor. These four typical airfoil sections should be sufficient to give an idea of the trends in design; thin symmetrical sections are desired.

ARRANGEMENT OF AIRFOILS. Arrangement of airfoils on a missile are governed by many factors, such as speed, range, launching period, and whether or not recovery of the missile is desired. Various combinations of the types shown to the left on the following page may be used. The sketches illustrate only the more common types of arrangements.



Common arrangements of airfoils



High speed configuration characteristics

In some arrangements, tail units are used as nose surfaces and some of the wing arrangements are used as tail units. The sketches to the right above show some of the typical configurations which may be used in conjunction with any of the foregoing arrangements to give good, high-speed performance. Note also the sketches of airfoil plan forms on the opposite page.

AERODYNAMIC LOADS. One of the problems of missile designing is aerodynamic loading. Such loading results from any change in the air-flow pattern. Aerodynamic loads are in general compensated for by control-surface positioning, sweepback, and fuselage configuration. Auxiliary equipment such as antennas, airspeed measuring devices, and various telemetering devices all add to the aerodynamic loads. All of these obstacles to an even airflow are being eliminated as new devices are made available to accomplish the same purpose.

In the transonic range, still more data is needed to fully determine the general behavior

of lift and drag on a missile and the basic laws governing this behavior. Systematic tests are still needed over the supersonic range to determine fully the general effects of various wing sections, airflow interference, and the combined-surfaces interference. But the design of missiles for supersonic flight has come a long way, and we can well expect continued improvement of guided missile design.

Aerodynamic Control

Aerodynamic control is the connecting link between the guidance system and the missile flight path. It is at this point where much regard is given to a smooth and exact operation of the control surfaces of the missile. The control must be powerful enough to produce the necessary change of direction. At the same time, it must have the best design configuration for the intended speed. Methods must be found for balancing controls and varying the control surfaces to meet the variations of lift and drag at different Mach numbers. The operating mechanism (or servo systems) must be made to operate readily to prevent instability in the missile.

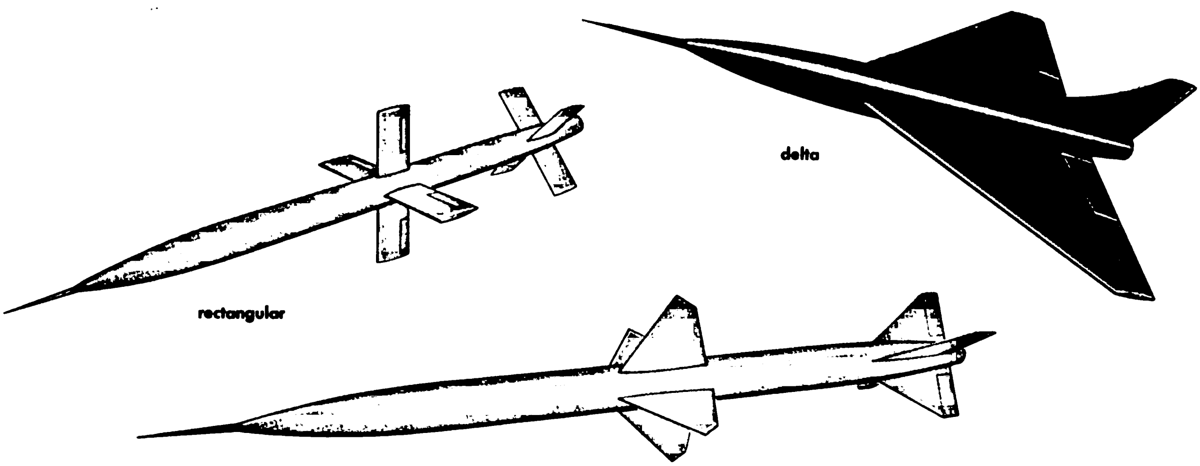
EXTERNAL CONTROL SURFACES. Fixed guide fins are the simplest type of control system for stability. The flight of a common arrow is an example of this type of stability, since the feathered fins at the rear of an arrow provide for a stable line of flight. This same principle of obtaining stability is used in missile design. Fixed guide fins are used in one way or another on all missiles. These fins are generally referred to as stabilizers of a specific type, such as horizontal stabilizer or vertical stabilizer.

In addition to fixed control surfaces, there are movable control surfaces referred to as flaps or movable vanes.

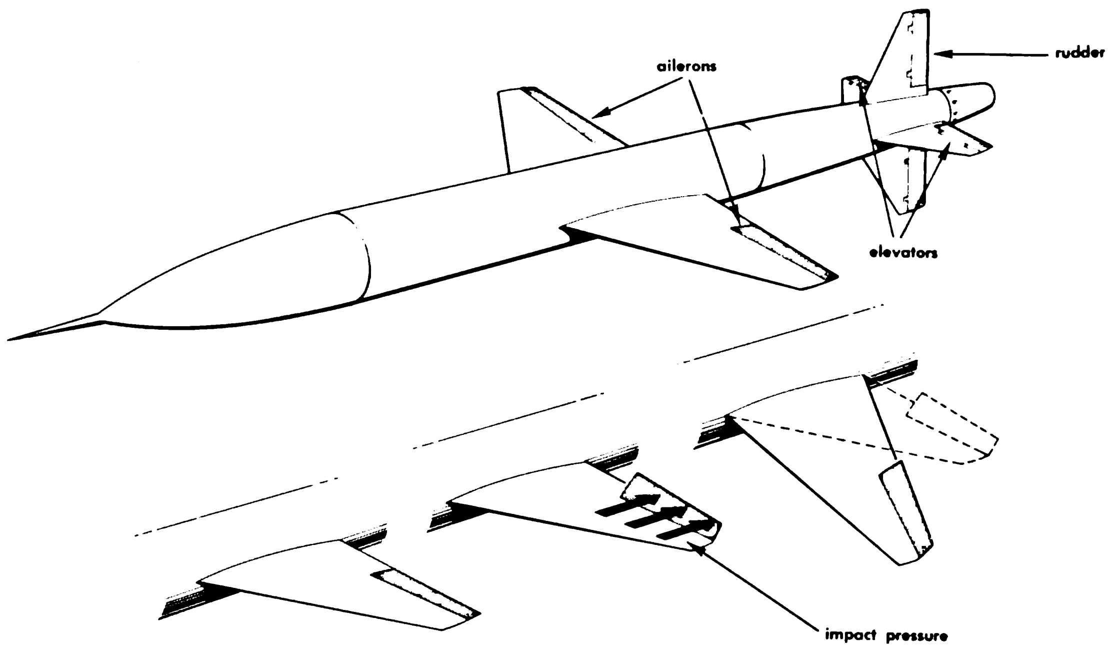
Keep in mind that a control surface is not effective until the airflow across the surface has attained sufficient speed to develop a force. As the speed is increased, the reaction to the control surfaces becomes more acute and often results in overcontrol.

The disadvantage of using guide fins without movable controls lies in the lack of the precise control necessary to follow a given course. To accurately control a missile, two general types of control surfaces are used. These types are primary and secondary controls. Primary control can be looked upon as the main controlling factor of the missile's path. This group alone could, under certain conditions, give satisfactory results. That is, if there were no unstabilizing conditions present, primary control could function satisfactorily unaided. However, by the use of secondary controls of various combinations which are discussed later, a missile can be controlled much more accurately and efficiently.

Primary Controls. Ailerons, elevators, and rudders are considered primary controls. A conventional *aileron* is attached to the outer



Airfoil plan forms



Primary control surfaces

trailing edge of the wings or main lifting surfaces. This conventional arrangement appears above. When one aileron is lowered, the opposite one is raised, thus controlling roll. They are coupled to the governing control system. Various control systems are taken up in later chapters.

Elevators are attached to the horizontal stabilizer on the tail. They are used for pitch control and are raised and lowered together.

Finally, a *rudder* is used to maintain directional control. It is attached to the vertical stabilizer and gives yaw control.

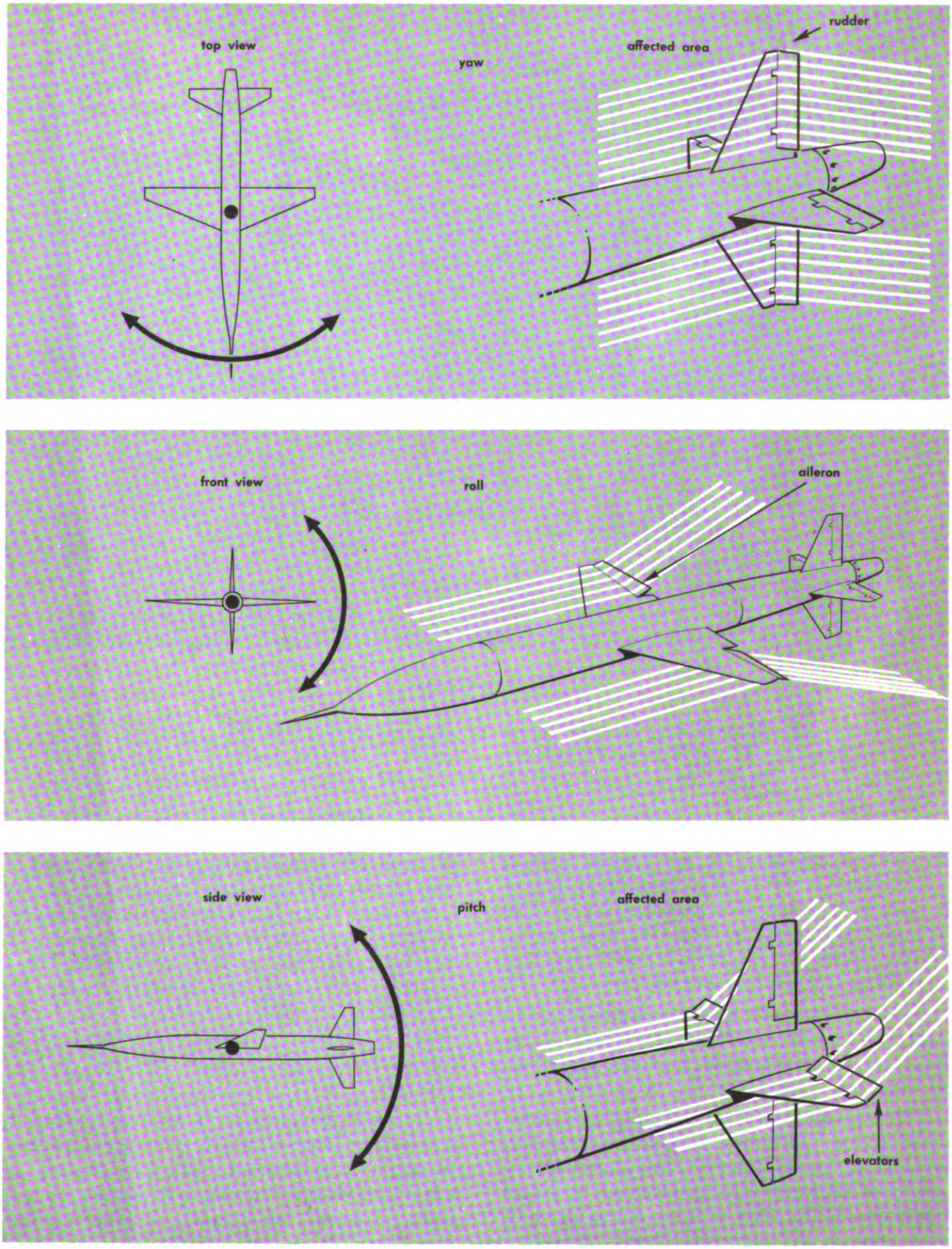
The illustration on the following page shows that the rudder, ailerons, and elevators control movement about the yaw, roll, and pitch axes respectively.

Control is attained by the previously mentioned control surfaces in that they present a surface to the existing air flow at an angle which will cause a force to exist. This force pushing against the control surface moves the wing or tail to which the control surface is attached in a direction opposite to the control-surface movement. Control surfaces along with the direction of movement resulting from the positioning of these surfaces are shown on page 38.

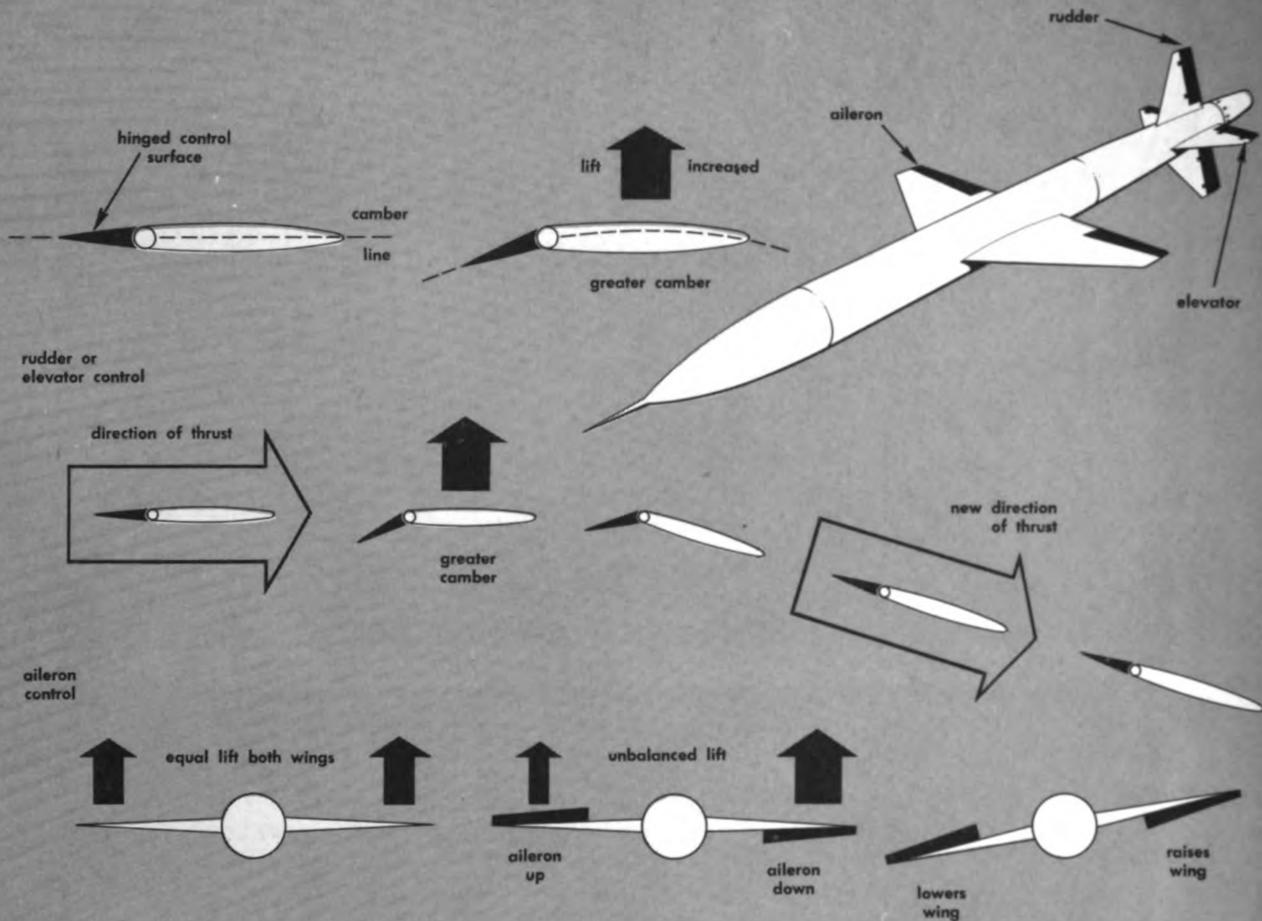
The wing, horizontal stabilizer, and vertical stabilizer could be considered a part of the primary control surfaces since they are the leading factors in determining the missile's path. But for our discussion of controls, we will regard the elements that tend to change the direction of travel as control surfaces.

Secondary Control Surfaces. In the secondary group of controls are tabs, spoilers, and slots.

Tabs can be divided into three types: fixed, trim, and booster. Their reaction in the relative air stream, however, is the same in all three cases. They always affect the control of a missile indirectly. They do not in themselves determine the direction of motion of the missile. For example, suppose it is desired to trim in pitch. To raise the nose of the missile the tab must be moved down. The primary control surface, in this case the elevator, is hinged to the horizontal stabilizer. The tab in turn is hinged to the elevator. A small movement of the tab on the trailing edge of the elevator causes a small force to be exerted on the primary control surface. The result of this force is a small movement of the elevator in the opposite direction; therefore, if the tab is moved down, the elevator moves *up*. Since the missile responds only to the primary control action, the tail is lowered,



Function of primary control surfaces



Control of flight

thus raising the nose of the missile. Note the above illustration. This function is exactly the same for directional trim (yaw) and for lateral trim (roll).

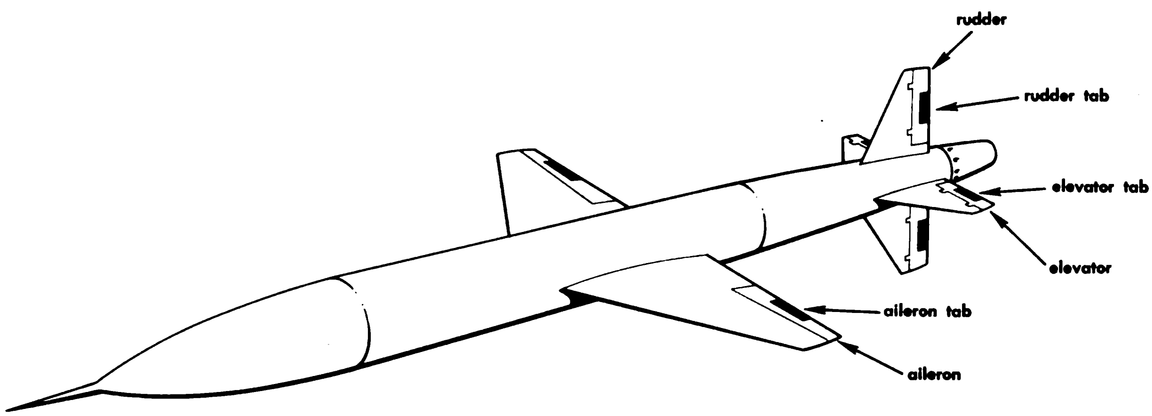
A fixed tab consists of a piece of metal attached permanently to the trailing edge of the primary control. It is bent uniformly in the required direction to trim the missile. The trimming is done in anticipation of a certain set of unbalanced conditions around the center of gravity at a given airspeed.

A trim tab is more complex in construction than a fixed tab, and it is controllable. Thus trim can be changed as attitude, speed, or altitude are varied.

A booster tab, sometimes known as a servo tab, is widely used where large primary control surfaces are used. The force and stress on the control system are greatly reduced by using the booster tab. This type of tab is directly connected to the control system. Therefore, when the system is actuated, the tab moves first. The tab movement in turn moves the primary control in the normal manner. The effect of the movement of the tab on the primary surface is considered instantaneous.

To summarize, a fixed tab is preset for a given condition of stability. A trim tab is controllable, and its setting can be changed at will over a wide range of conditions. A booster tab is used to assist in moving control surfaces of large area. Since the performance and design of these controls is well understood, they probably will be used on missiles in one form or another for sometime to come. Even so, in the case of supersonic missiles, new systems of aerodynamic control may have to be developed.

As aircraft speeds were increased closer to the speed of sound, the performance of surface controls became more and more critical. A surface control is nothing more than a hinged flap, or airfoil, attached to a larger airfoil which may be, as previously stated, a wing or a stabilizer. Our engineers soon found that, at speeds over the 300-mph range, peculiar *vi-*



Location of tabs

brations called *flutter* occurred on controls, especially in the case of ailerons. These vibrations became more severe as speed increased until in some cases they reached such proportions that surface controls disintegrated and sometimes caused failure of the entire wing or tail structure. This has been overcome to a great extent by increasing the rigidity and strength of the primary controls and by balancing these surface controls statically and dynamically.

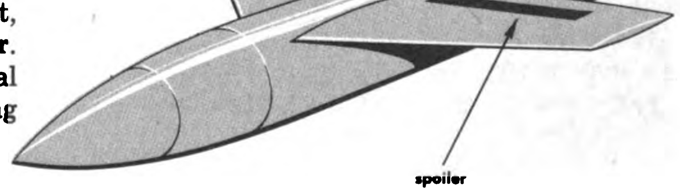
A method of overcoming bending on a wing is by the use of *spoilers*. They may be of one solid unit or a series of units that tend to interrupt the negative lift on a wing. Spoilers are *recessed* into the upper camber of the wings.

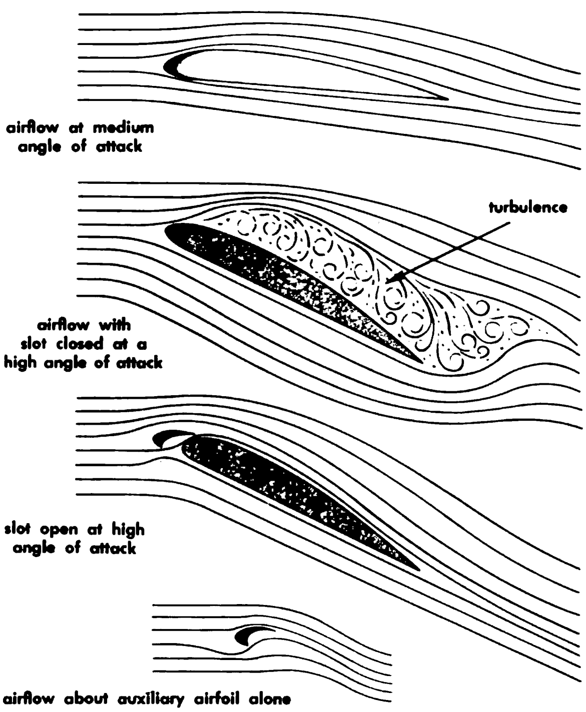
The spoiler in the illustration on the right consists of a solid hinged flap. When the spoiler is not used, the flow of air over the wing is smooth and uninterrupted; thus the full lifting power of the wing is realized. However, assume that a gust of air has caused the left wing to drop. The control system instantly calls for the spoiler on the right wing to extend. As the spoiler extends, the negative lift pattern on the right wing is "spoiled," or reduced a considerable amount, by the turbulence created by the spoiler. The wings then tend to return to the original position. A spoiler may cause more drag than the conventional aileron; therefore, when the spoiler is extended, a yawing moment may be initiated strong enough to cause the rudder to be actuated simultaneously with roll control.

A *slot* is basically a high-lift device and is located along the leading edge of the wing. In the region of normal angle of attack, the slot is ineffective. But as shown on the following page, when a missile reaches high angles of attack, the slot can be opened to allow air to spill through. A slot performs a function exactly opposite to that of a spoiler. It is a valuable device for increasing lateral stability, landing high-speed aircraft, and for preventing stalls in slow aircraft.

One type of slot operates automatically with an increased pressure difference resulting from high angles of attack. A disadvan-

Location of spoilers on a wing surface





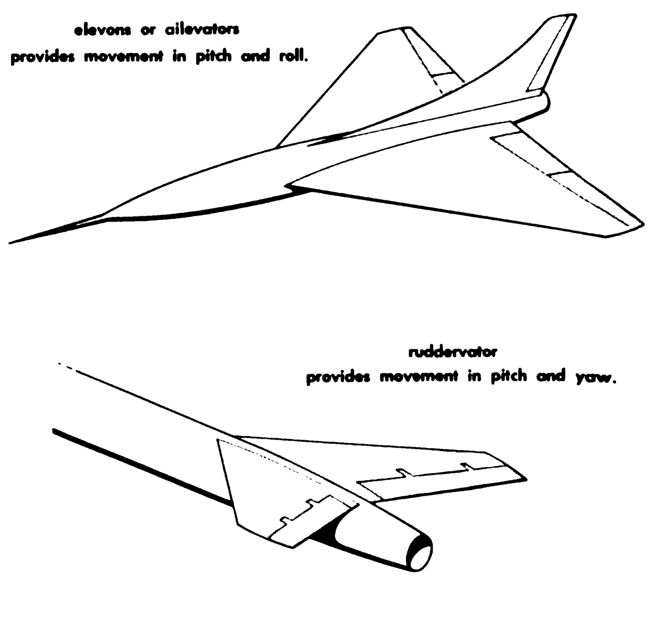
Airflow with slot

tage of this mechanism is that it must be kept locked until needed because maneuvers at high speed calling for large angles of attack would automatically open it, causing serious reduction in performance or even structural failure of the slot.

Dual-Purpose Controls. Up to this point the controls discussed have been of a conventional type. As missile speeds increased, new control surfaces were developed. New types of primary controls are elevons, ruddervators, and ailevators. As the names indicate, they consist of control surfaces that accomplish two purposes. For example, an elevon takes the place of an elevator and an aileron, giving control of pitch and roll. An ailevator is the same as an elevon. Ruddervators are used for yaw and pitch control. Observe the illustration at the right.

Variable Incidence Control. Use is sometimes made of a variable incidence control to overcome the problem of flutter and the need for structural strength of control surfaces and yet have a control that is sensitive and effective at various speed ranges. This type of control varies the positioning of an entire airfoil section rather than just part of it. This method eliminates the disturbing factors that are set up between two surfaces, such as the surfaces of a slot and a wing, which make an angle with each other. At high speeds, this type of control surface is especially well adapted because little control movement is required to displace the missile a considerable distance. The variable incidence control can be used on the wing, horizontal stabilizer, or vertical stabilizer.

Canard Structure. A canard structure consists of a stabilizing plane and elevator control which is located ahead of the center of gravity, or wing and lifting surfaces. During the early development of the airplane it was found to be easier from the standpoint of construction to install the tail plane at the aft end of the fuselage. This trend has been followed throughout the years. However, the Canard adaption to missiles is now in the experimental stages by some manufacturers. The forward wing may consist of a stabilizing plane with a surface control attached to the trailing edge, or it may consist of a pivoted wing in which case the entire stabilizing plane has a movable angle of incidence (variable incidence control), either positive or negative.



Dual-purpose control surfaces

Dive Brakes. Dive brakes may be of two types: one that uses a *single hinged* flap that works as a resisting surface or one that uses a *split surface* for deflecting the air stream. Dive brakes can be considered as part of the primary control surfaces. As the name implies, they are used to give a braking effect by creating an excessive drag on a missile, thus slowing it down.

CONTROL AT STARTING SPEEDS. Until a speed at which the airfoil sections have an aerodynamic stabilizing effect is attained, there must be a means of control other than the external control surfaces. This means may be supplied by the use of exhaust vanes or jet control.

Exhaust Vanes. Exhaust vanes are surfaces which are installed directly in the exhaust path of the jet engine. When the position of the vane is changed, it deflects the exhaust and causes the thrust to be directed in opposition to-the exhaust vane. The velocities of the exhaust are sufficient to give adequate control until the time the missile has attained speed enough for the external controls to take over. Because of the tremendous heat in the exhaust, the life of exhaust vanes is generally short.

Jet Control. Various systems of jet control, illustrated on the following page, are being developed for supersonic flight. One means of jet control is accomplished by changing the position of the jet engine in such a manner as to give the desired direction of motion. This method utilizes the gimbaled engine. Two serious objections to the method are that all of the various fuel lines must be made flexible, and the control system that actuates the jet must be extremely strong.

An easier system of jet control is accomplished by placing several jets at various points about the missile body. Control is accomplished by using one or another of these jets as desired thus giving different directions of thrust. This method eliminates use of the outside control surfaces, affording a cleaner missile surface. The method may be referred to as deflection charge control.

Air Deflector Control. Another means of control is the use of a deflector at the intake of the jet. The deflector modifies the air flow pattern about the missile, thus changing the direction of the missile as a result of drag. However, this method is rarely used.

SPECIFIC CONTROL APPLICATIONS TO GUIDED MISSILES. As you know, the standard method for control of aircraft around the vertical axis is the rudder located in the tail assembly. Likewise, the control around the lateral axis is performed by the elevators located in the tail assembly. And the control of roll around the longitudinal axis is performed by the ailerons attached at the trailing edge of the outer wing panel as near the wing tip as possible.

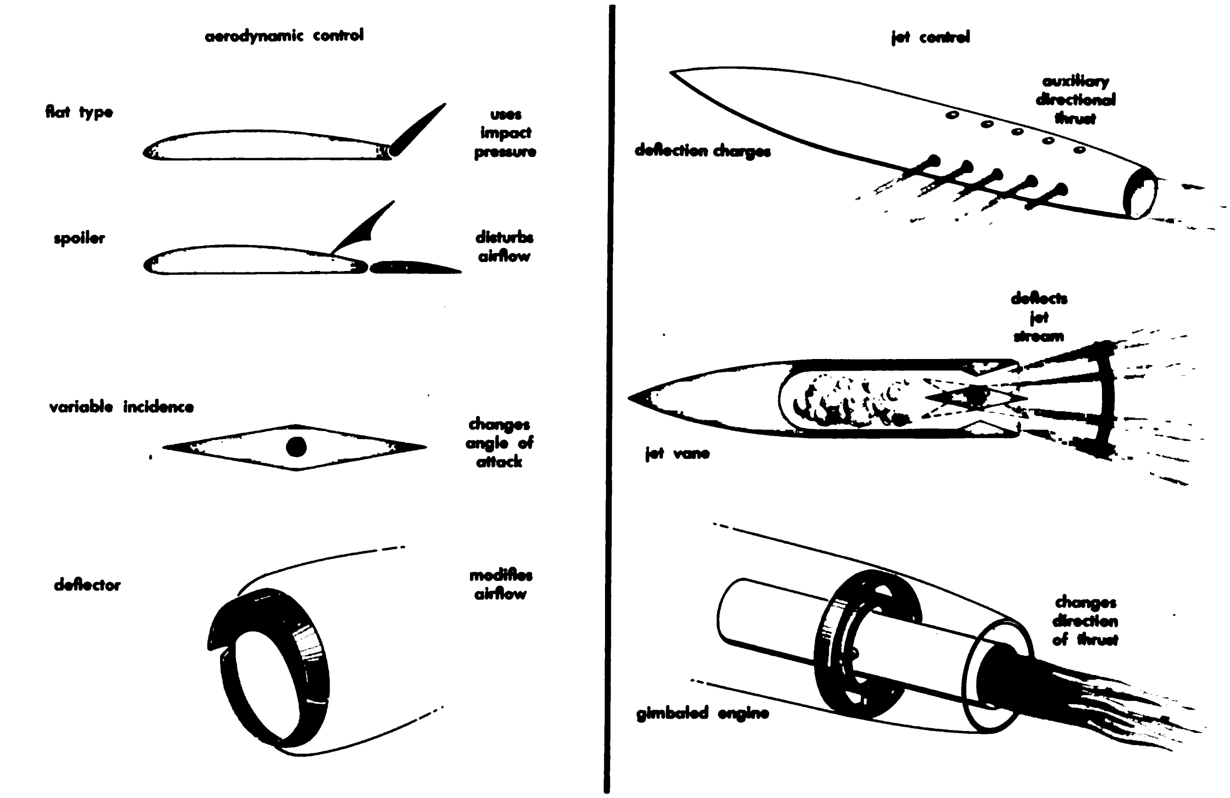
Because of the high speed of guided missiles, it generally is necessary to control a missile by controls other than the standard methods. For example, in one configuration the tail assembly is fixed. Directional control is performed by flaps located at the trailing edge of vertical control surfaces (see bottom of page 43). When the flaps are moved from neutral position, the chord of this vertical wing is effectively changed to a given angle of attack, depending upon angular displacement of flaps. Changing the angle of attack creates a *differential* in pressures on the wing, thus affecting lift. This results in the missile turning about its vertical axis, and a change of direction takes place.

Correction for roll is performed by ailerons located in the wing tips of the vertical wings. Channel guides hold the ailerons at a fixed angle of incidence with respect to the wing panel chard. When the aileron is extended, the positive angle of attack exerts an aerodynamic lift which tends to correct an error in roll.

The change in attitude of the missile around its lateral axis is accomplished by means of flaps on its horizontal wings. When the flaps are lowered, an increase in lift tends to lift the nose of the missile, and conversely, the raising of the flaps lowers the nose.

Stability of Missiles

Since stability of a missile has a direct effect on the behavior of the controls, a high degree of stability must be maintained. As missile speeds are increased there are definite stability changes caused by center-of-pressure shifts.



Steering methods

A pressure shift causes variation in the flows that are acting on the surfaces of a missile. Even in pure supersonic flow, variations of the Mach number cause center-of-pressure shifts. Let's consider an example. Referring to the top illustration on the next page, note that the missile is in level flight. It is longitudinally stable about its lateral axis through the center of gravity. Air flow over the wing is deflected downward toward the elevator. This angle of deflection is called the *downwash angle*. When lift decreases as a result of reduced speed this downwash angle decreases resulting in pressure changes. There will be points where unstable conditions are set up as a result of such pressure shifts resulting from different airflow patterns. Once an uncontrollable condition is set up, it must be compensated for by the control surfaces or by quickly changing the speed to a point of stability.

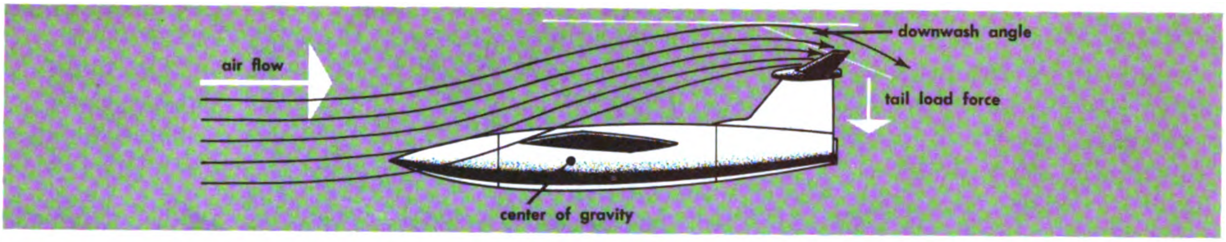
As we learned earlier, unstable conditions are dominant at the transonic speeds. Most missiles have some dive control and roll recovery devices to overcome unstable conditions. An example of such a device is an arrangement in which horizontal tail surfaces are placed high on the fin to help overcome downwash conditions.

Ailerons and rudders come in for their share of trouble. The unstable flow over the wings causes the ailerons to oscillate, creating a condition known as "buzz." A similar condition called "snaking" may exist about the directional axis. A part of these troubles may be compensated for by nonreversible control systems or by variable incidence control surfaces.

New problems of stability are constantly arising as speeds and altitudes are increased. Close cooperation between aerodynamicist and manufacturer is necessary to solve these problems.

Propulsion Aerodynamics

Propulsion aerodynamics includes such items as the design of low-loss air intakes for ramjets; design of subsonic diffusers, nozzle,



Downwash

and tail pipe; and the effects of jet exhaust on drag and flow about the body of a missile. Aerodynamic-mechanical problems of fuel control, regulation of the proper amount of air intake, and the stability of the internal flow and combustion process must also be dealt with. As new developments are made in propulsion systems, control by way of the jet stream may be utilized to a greater extent, eliminating the requirements for external control surfaces. This elimination of wing surfaces is prominent today on guided missiles designed to follow a ballistic type of trajectory.

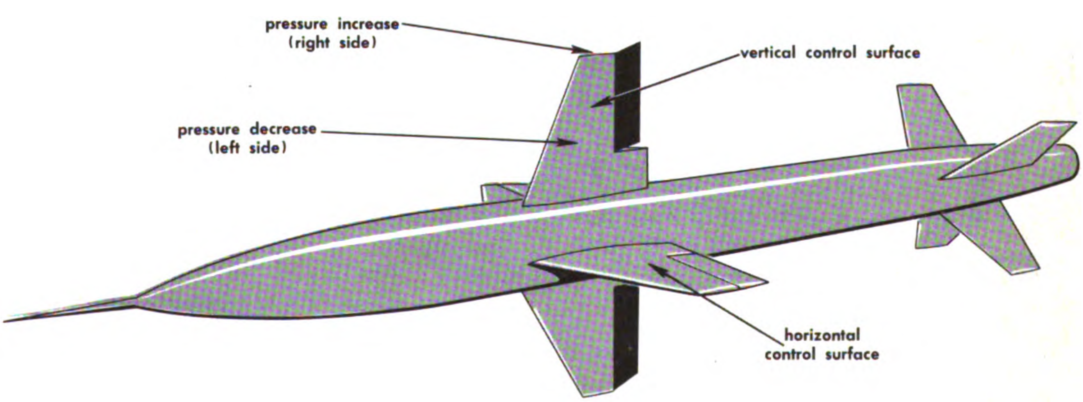
Even though aerodynamic improvement is worked for constantly, present flight achievement is meeting great difficulty in obtaining workable configurations which are capable of consistent supersonic performance. As yet, aerodynamic design is behind the advancement of propulsion systems, now capable of delivering tremendous thrust.

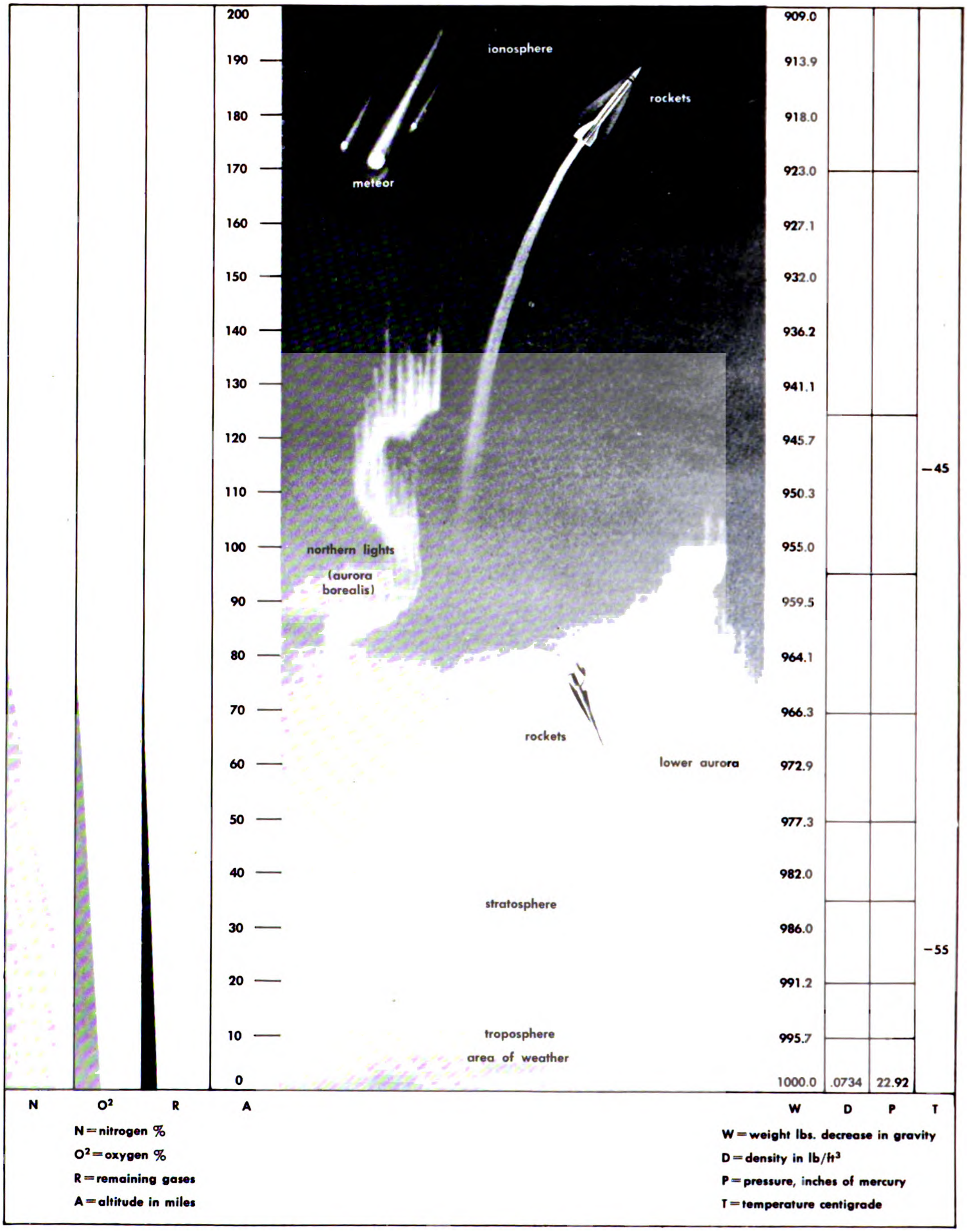
Effects of the Atmosphere on Flight

Because of the greater density of air at low altitudes, considerable resistance to high-speed travel of a missile is developed. As a result, the jet propulsion power plant is not efficient at these altitudes because the major part of the available thrust is used to overcome drag rather than for acceleration. Drag on a missile varies as the square of the velocity of the missile. Once the missile has reached its maximum speed, the value of the thrust developed is equal to the drag value.

At high altitudes, low-density conditions prevail, making increases in speeds possible because thrust is used more for acceleration and less to overcome drag. At low altitudes, the jet power plant supplies energy to overcome air resistance and gravity and to accelerate the missile, while at higher altitudes the powerplant supplies the energy needed to overcome gravity and accelerate the craft. The chart on the following page shows pres-

Directional control using flaps in the vertical control surface





Characteristics of the atmosphere

sure, temperature, and density changes of air that occur with increase in altitude.

The stratosphere is considered to be the region in which all long-range missile flights will be made. The stratosphere, located outside the limits of the troposphere, is a region of contant temperature and little air convection. By carrying out flights in the stratosphere, the advantages of low drag, high speed, low fuel consumption, and greater range are obtained. The disadvantages of flight in this region are the low temperatures that prevail and the scarcity of oxygen.

Techniques for Studying High-Speed Aerodynamics

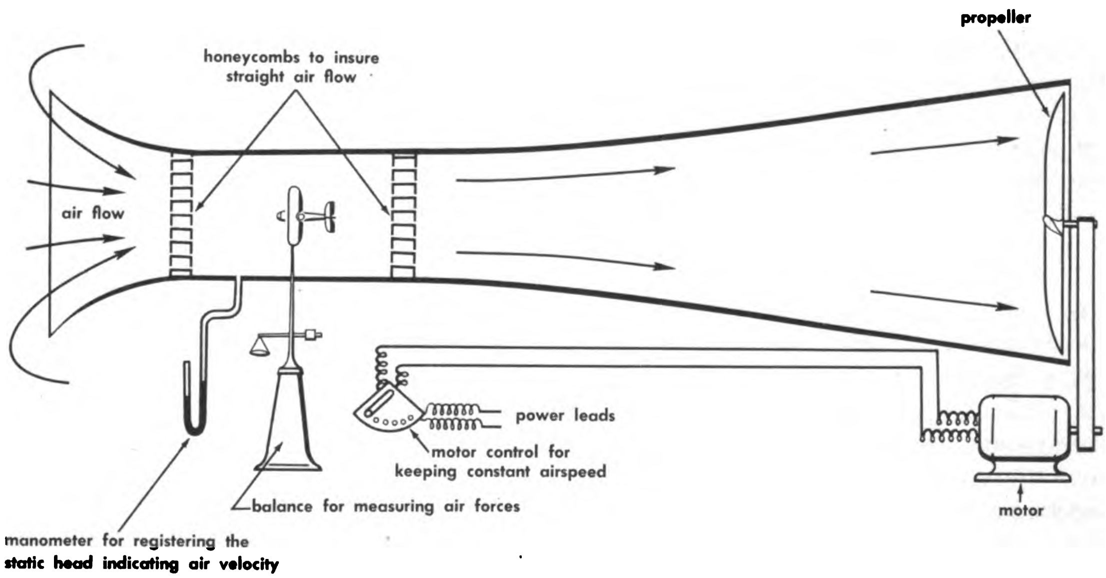
Techniques for studying high-speed aerodynamics are one of the major problems of supersonic aerodynamics. Developing missiles for supersonic speeds requires special procedures and special equipment. New wind-tunnel designs are helping to meet this problem.

WIND TUNNELS. While wind tunnels for some time have been used with great success in the subsonic field, it has been necessary to improve and enlarge them to meet the needs of missile study.

As soon as a missile reaches Mach 1, the effects of shock waves must be considered. In setting up such velocities as this in wind tunnels, it was found the shock waves are reflected off of the tunnel walls. Therefore, the diameter of the tunnel compared to that of the model must be very large in order to approach the conditions of actual flight.

New techniques in wind-tunnel design have eliminated the reflection of the waves, or "choking" effect. The problem of supersonic tunnel design was not in many ways as great as that of subsonic tunnel design, but a bottleneck in progress developed because of transonic speeds.

Wind-tunnel choking, scale effect, or Reynolds number, which up to this time had been well understood, took on new aspects in the transonic region. Results of scale-model tests could no longer be truly reliable; consequently, it was necessary to develop larger supersonic wind tunnels. Wind tunnels have been constructed which accommodate complete missiles of the smaller sizes. Most of the larger missiles are tested by sections, using full-size sections. The accompanying figure shows the basic layout of a wind tunnel:



Wind tunnel construction

REPORT OF WIND-TUNNEL TESTS. The following are parts of reports made in wind-tunnel tests:

> . . . Eight data runs were made at Mach number 2.23 to measure the turbulent-boundary-layer shear stress along a flat plate, and to obtain total pressure surveys through the boundary layer. The plate was tested at angles of attack of 0, −1.25, and −3.00 degrees in order to vary the effective free-stream Mach number . . .
>
> . . . Thirty-five data runs were made at Mach number 1.73 to obtain the stability, control, and drag characteristics of composite configurations of the 1/82.5 scale model. The model was mounted with −24 roll-indexing balance to record six-component data between −8 and +12 degrees indicated angle of attack at 0 and 90 degrees roll attitude. Eleven surface deflections of 0, −6, and −10 degrees were tested. . .

These samples give you some idea of what is included in the thousands of tests that must be made in the development of a missile.

While missile testing has improved considerably in recent years, no doubt additional techniques of testing will become necessary in the future.

Progress Reports on Guided Missiles

During the development of any missile, periodic reports are published giving new developments and the results from tests made on current models. Parts of such reports are given on this and the next page. The examples are given so that you may realize the complexity of the problems that confront designers in the development of new missiles.

The following exemplifies a typical progress report on a missile, showing what the related influencing factors are:

> . . . The tactical requirements of this guided missile made it necessary that the tactical version of the missile be capable of carrying a warhead of 2000 pounds and have a range of approximately 1000 miles. It was also desired that the tactical version would be in production in the shortest time possible. This made it advisable to use standard and proven design features. In order to use a reliable and proven power plant, a low supersonic speed was chosen that allowed the use of a conventional turbojet engine.
>
> The cruising speed was chosen as Mach 0.85. This speed helped in making the vulnerability less. An added reduction in vulnerability was realized by selecting a high cruising altitude of 60,000 feet. High wing loading made it impractical to choose a cruising altitude above 60,000 feet.
>
> Since this missile was to attain supersonic speed during the terminal dive, it was necessary that it exhibit good stability and control characteristics in the transonic and supersonic regions. These high speed requirements also called for clean design.
>
> To have a clean design, it was decided to have a symmetrical configuration throughout the missile. This symmetrical design in the transonic region of speed is desirable because one advantage for transonic speeds occurs at the zero-lift dive condition.
>
> A sharp sweepback angle and low aspect ratio was also decided upon in order to improve the Mach speed characteristics. The airfoil sections were made thin and symmetrical; this also aided the high-speed performance. Having a swept-back wing made it possible to eliminate the horizontal tail surface since longitudinal control could be obtained by the trailing edges of the wing. Ailevators provided the necessary lateral and longitudinal control. Elimination of a horizontal tail surface greatly simplified the design and eliminated air-flow problems that would otherwise be present. It also eliminated other problems that would have been met in designing separate control surfaces to have the same critical Mach number. The ailevators were placed near the fuselage which reduced the length of the connections to the actuators within the fuselage.
>
> The design of the vertical tail section was determined by the same considerations as those for the wing sections. Good

directional stability was attained by proper location of the single vertical tail section. The use of a double vertical tail had been considered, but since it was desirable to provide a landing gear for recovery, the double tail was abandoned. Since the cost of flight-testing the missile was so great it was worthwhile to provide this landing gear.

Control response was investigated at the cruising speed to be sure that the missile had the required maneuverability. It was decided to use banked turns since the skid-turn performance was inadequate. This change caused some further complications within the autopilot and controls system . . .

The following discussion might be typical of another progress report of a considerably smaller missile than the one discussed above:

. . . In this particular missile, the size and shape was the main consideration; therefore, the main attention was given to the number, size, and location of control surfaces. Designing a missile according to specified size and shape gives rise to particular aerodynamic problems. The factors of greatest importance are those concerning drag, stability, and maneuverability.

It was desirable that this missile exhibit good stability at low speeds and in adverse wind conditions. A highly swept-back, delta wing was chosen to give the desired lift and stability at low speed and at the desired upper limit of speed of Mach 3. Since lifting surfaces suffer a loss in lift at higher speeds, the chosen wing showed good performance under all conditions.

To limit the overall length of the missile, it was decided to eliminate a tail section and use four symmetrically placed wing surfaces. Modifications of the control system were made to function with the planned configuration.

To attain the desired maneuverability of this missile, having a small lifting surface, it was decided to use adjustable jet vanes in addition to the wing controls. These jet vanes were located in the motor exhaust and coupled directly to the control surfaces. The jet vanes provided good control at low speeds before the wing surfaces became aerodynamically effective. These jet vanes were graphite coated to withstand the high temperatures of the motor exhaust . . .

From the discussions above, you can readily see that considerations given to missile design are many and varied. You can see that no one problem stands alone when a new missile is being developed.

TRENDS IN CONSTRUCTION OF MISSILES

Up to the present time, fuselage construction of missiles seems to be following the usual pattern of semimonocoque or full monocoque design. A semimonocoque type usually has four main longitudinal members with secondary members attached to them. In full monocoque, the construction depends entirely on the skin being attached to the secondary members, thus forming a shell-like structure.

A new metal under consideration for covering of structures is titanium. This metal is about 60% heavier than aluminum but only about half the weight of steel. Its alloys are several times stronger than aluminum and rival the best steel alloys. Titanium may eventually be used to replace aluminum for skin-stressed structures as well as nonstructural parts. Its principle drawback at the present time is its extremely high price in sheet form.

Another important characteristic of titanium is its resistance to high temperature. The problem of surface temperatures has been mentioned before in regard to the structural problems. Apart from this, the temperature rise presents a serious problem of premature explosions of warheads, combustion in the fuel lines, and damage to operating equipment. Titanium can be used in the 300° to 800° temperature range encountered in the transonic and supersonic speed ranges.

In practically all cases, missiles designed for combat are not intended for recovery; therefore, there is no need for them to have landing gear. However, during research and development stages, it is economical to install landing gear as a means of recovering the

missile. It is also essential to recover missile which are to be used as reconnaissance vehicles.

There are other means of recovering a missile besides the use of landing gear. A parachute can be used to lower guided missiles to the earth with little damage. Another device for recovering a missile is a long pointed shaft attached to its nose. The shaft tends to absorb the impact of the landing. This landing procedure is quite limited in application.

The most satisfactory means of missile recovery, however, is by the use of some type of landing-gear arrangement. The landing gear may be wheels or some type of skid.

Whenever landing gear is incorporated into a missile, it gives rise to other problems of design. Adequate space must be provided in the interior of the missile for the mechanism. Stability characteristics change considerably, and allowances must be made for any air-flow pattern changes that may occur as a result of adding the landing gear.

As already pointed out, most combat missiles do not have landing gear or any other means of recovery since they are "one-shot" vehicles. Any available space is used for control equipment, fuel, or warhead material.

Here, again, in the structural aspects of missile development we can expect continued advancement. New materials and new methods of construction will continue to increase the efficiency of missiles.

YOU AND AERODYNAMICS

In this chapter you have been given a general picture of the problems involved in the field of aerodynamics. As you can see, it would be beyond the scope of any one text to cover every phase of aerodynamics in complete detail.

In your work with a missile squadron, you won't need a complete knowledge of aerodynamics. But you will find it to your advantage to have an understanding of the basic concepts of aerodynamics presented here.

Chapter 3

Propulsion of Guided Missiles

Guided missiles must move with a high velocity so as to lessen the probability of interception and destruction by enemy countermeasures. Missiles must also be capable of intercepting and destroying high-velocity enemy missiles and manned aircraft in flight. Until recently the jet engine was the principal power plant used in propelling aerial vehicles at supersonic speed.

Until the start of World War II, the reciprocating engine-propeller combination was quite satisfactory as a propulsion device for aircraft. However, as speeds increased, the combination proved unsatisfactory because propeller thrust diminished rapidly after a certain speed was reached. This condition necessitated extremely large engines to produce sufficient horsepower to give any further increase in speed. In addition, when approaching the speed of sound, shock-wave formations on the propeller effectively reduced the thrust. Even though investigations into propeller designs indicate that much of their limitations may be overcome, it is necessary at the present time to use jet propulsion for high subsonic and supersonic speeds.

BASIC FORMULAS AND LAWS UNDERLYING JET PROPULSION

> "Jet propulsion is a means of locomotion brought about by the momentum of matter ejected from within the propelled body."

Because of this definition of jet propulsion, you may sometimes see jet-propulsion engines referred to as reaction-type engines. This label is not sufficiently specific, however, since any body moving in a fluid works on the reaction principle if it is self-propelled. For instance, the action of a conventional propeller consists of increasing the momentum of the air, and the propeller thrust is the resultant reaction.

The ordinary propeller-driven aircraft is not a form of jet propulsion because the working fluid is not ejected from within the vehicle. If the propeller were ducted and the air allowed to pass through the vehicle, then the vehicle would have mechanical jet propulsion. None of the jet-propulsion engines of missiles utilize the mechanical method.

In understanding the principles involved in jet-propulsion systems, it is important to consider a basic mathematical analysis of the manner in which the thrust of a jet engine is developed. Remember that thrust is any force tending to produce motion in a body or alter the motion of a body.

Newton's second and third laws of motion are the underlying principles on which jet propulsion is based. Newton's second law of motion tells us that when an *unbalanced force* acts upon a body it causes the body to *accelerate* in the direction of the force. The acceleration produced is *directly proportional* to the unbalanced force and *inversely proportional* to the mass of the body. Thus, the formula for force is:

$$F = Ma, \tag{1}$$

where "F" is the unbalanced force in pounds, "M" is the mass of the body in slugs, and "a" is the acceleration produced in feet per second per second.

For any given mass the gravitational attraction of the earth on that mass varies with geographical position or distance from the earth's center. This attractive force exhibited by the earth is called weight. As a result of this variation in the weight of matter, it is necessary to use the following representation of mass in your mathematical expressions. That is:

$$M = \frac{W}{g}, \tag{2}$$

where "M" is mass in slugs, "W" is the weight in pounds, and "g" the acceleration produced by the earth's gravitational attraction. Both "W" and "g" vary in such a manner as to keep ' M" constant.

By definition, acceleration "a" is the rate of change of velocity "v" and may be expressed as:

$$a = \frac{v_2 - v_1,}{t} \tag{3}$$

where "v_1" is the initial velocity of some mass, "v_2" is the final velocity of that same mass, and t is the time required to change the velocity from v_1 to v_2. You may now rewrite the original equation, F=Ma, as:

$$F = \frac{Mv_2 - Mv_1}{t} \tag{4}$$

Since by definition, "Mv" is called momentum, it can be stated that the thrust of a jet engine is equal to the rate of change of momentum of the working fluid. The above equation is frequently written:

$$F = m\ (v_2 - v_1), \tag{5}$$

where "m" represents "M/t" and is called the mass rate of flow of the working fluid in slugs per second. Now, by substituting equation 2 in equation 1, Newton's second law may be represented by:

$$F = \frac{W}{g}a \tag{6}$$

In application of this equation to jet propulsion, "F" is the unbalanced force which accelerates the working fluid through the exhaust nozzle, and "a" is the acceleration in ft/sec^2. In accordance with Newton's third law of motion, the forward thrust (T) of a jet-propulsion unit is equal and opposite to this unbalanced force (F).

Now let "W" equal the total weight in pounds of working fluid that flows through the unit during the thrust-producing portion of operation, and let "t" equal the total time required. Then, "W/t" will equal the weight rate of flow in pounds per second. Letting w=W/t, you can now write an equation for thrust:

$$T = \frac{w}{g}(v_2 - v_1), \tag{7}$$

where: T=Thrust in pounds

w=Weight rate of flow of working fluid in pounds per second

v_1=Intake or initial velocity of working fluid

v_2=Exhaust or final velocity of working fluid

g=Acceleration of gravity (taken as 32.2 ft/sec^2)

While the above equation is actually the method for calculating the force exerted upon the working fluid, it also gives you the value

of the thrust forcing the jet engine forward through space.

Examples of Newton's second law are shown in the illustration below. The first figure shows a body weighing 32.2 lbs. subjected to a force of 20 lbs resulting in an acceleration of 20 ft/sec². The next figure shows a body weighing twice as much as the previous one but acted upon by a like force of 20 lbs. Note that the acceleration imparted to the body in this case is only 10 ft/sec². The third figure again shows a body weighing 32.2 lbs, but the applied force is decreased to 10 lbs. Note that the acceleration has also decreased as compared to the first illustration.

$$F = \frac{W}{g} a$$

$$20 \text{ lbs} = \frac{32.2 \text{ lbs}}{32.2 \text{ ft/sec}^2} a$$

$$20 = 1a$$

$$a = 20 \text{ ft/sec}^2$$

$$F = \frac{W}{g} a$$

$$20 \text{ lbs} = \frac{64.4 \text{ lbs}}{32.2 \text{ ft/sec}^2} a$$

$$20 = 2a$$

$$a = 10 \text{ ft/sec}^2$$

$$F = \frac{W}{g} a$$

$$10 \text{ lbs} = \frac{32.2 \text{ lbs}}{32.2 \text{ ft/sec}^2} a$$

$$10 = 1a$$

$$a = 10 \text{ ft/sec}^2$$

In the drawing of a jet engine on the following page, the applied force results from the chemical reaction between the air and the fuel. Observe that the resulting acceleration given to the weight of working fluid is in the change in velocity of the working fluid from v_1 to v_2. The thrust that drives the engine forward is equal in magnitude but opposite in direction to the force that drives the working fluid rearward. The drawing of the jet engine on next page illustrates a typical situation in which an engine develops thrust in accordance with Newton's second law of motion.

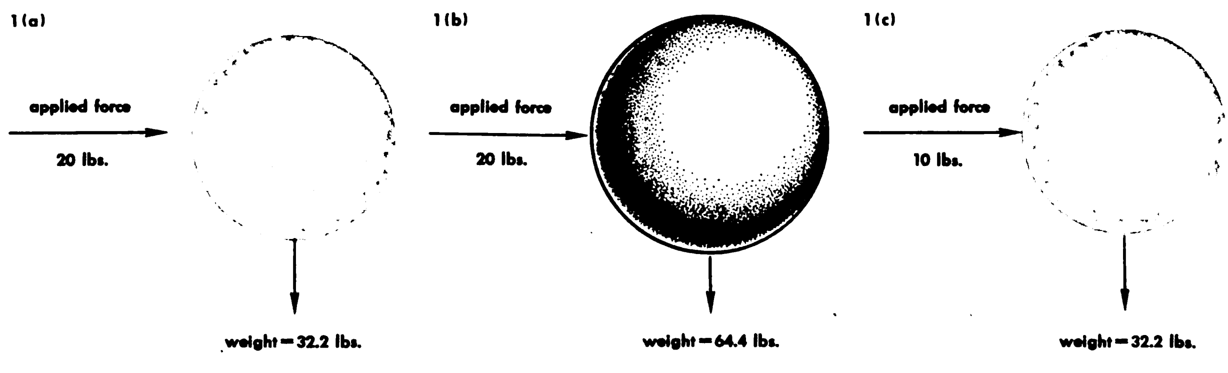
In the case of a rocket-propulsion engine where the working fluid, or fuel, is stored within the unit, the original velocity (v_1) is *zero*. Thus, the thrust formula for a rocket is:

$$T = \frac{w}{g} v_2 \qquad (8)$$

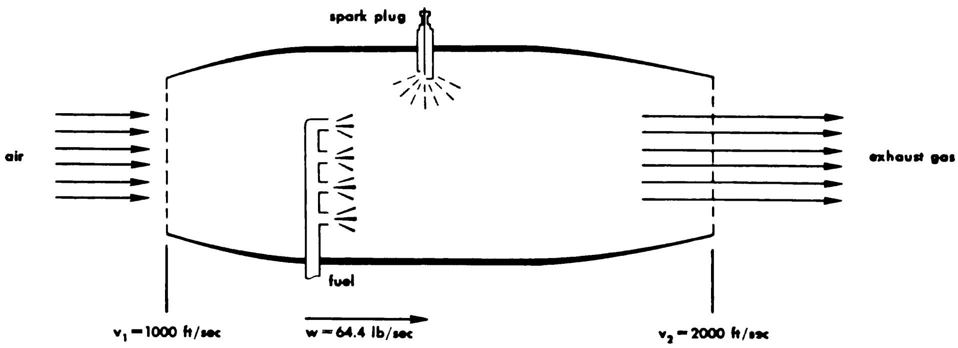
Notice that thrust is always expressed as *pounds of force* and *not* in terms of work or horsepower. A jet engine that is being static fired is not moving through any distance and is therefore doing no work, nor is it developing any horsepower. You can, however, calculate the thrust horsepower for a moving missile powered by a jet engine, by the following relationship:

$$\text{Thrust Horsepower} = \frac{\text{Speed of missile (mph)} \times \text{Thrust (pounds)}}{\text{One Horsepower (mile-pounds per hour)}} \qquad (9)$$

The *horsepower* in the denominator is equivalent to 375 *mile-pounds per hour* and is derived from the more frequently used horsepower unit of 33,000 foot-pounds per minute. Thus, a missile of the V-2 type, if developing



Examples of Newton's second law of motion



$$\text{thrust} = \frac{w}{g}(V_2 - V_1)$$

$$\text{thrust} = \frac{64.4 \text{ lb/sec}}{32.2 \text{ ft/sec}^2}(2000 \text{ ft/sec} - 1000 \text{ ft/sec})$$

thrust = 2000 lbs.

Practical example of Newton's second law of motion

56,000 pounds of thrust and traveling at 3750 miles per hour, would develop 560,000 horsepower. Using equation 9, we have:

$$\frac{3750 \text{ mph} \times 56{,}000 \text{ lbs}}{375 \text{ mi-lb/hr}} = 560{,}000 \text{ hp} \qquad (10)$$

It is important that you remember, however, that jet-propulsion engines are always rated in terms of pounds of thrust rather than in horsepower.

A common misconception in the case of jet engines, especially the rocket, is that the exhaust gases impart thrust to the missile by pushing against the outside air. The air in no way helps drive the rocket. It acts only to impede the rocket's motion by building up resistance and hindering the rapid straight-line ejection of the exhaust gases.

Consider the thrust on a rocket-powered missile as being composed of the sum of two terms. The first term, the *momentum thrust,* is the product of the *mass rate of flow* of the propellant and the *exhaust velocity* relative to the vehicle. The second term, the *pressure thrust,* consists of the product of a *cross-sectional area* of the exhaust jet leaving the vehicle and the difference between the *exhaust pressure* and the *atmospheric pressure.* On this basis the fundamental rocket equation for thrust may be stated as follows:

$$T = \frac{w}{g} v_e + (P_e - P_a) A_e, \qquad (11)$$

where: T = thrust developed

w = weight rate of flow of propellants

g = acceleration due to gravity

v_e = velocity of the gases at the exit

P_e = pressure of the gases at the exit

P_a = atmospheric pressure

A_e = cross-sectional area of the exhaust jet

When the atmospheric pressure is equal to the exhaust pressure, the pressure-thrust term is zero, and the thrust is expressed as:

$$T = \frac{w}{g} v_e \qquad (12)$$

This condition represents maximum thrust for a given propellant and chamber pressure. A rocket-motor design which permits the expansion of the propellant products to the pressure of the surrounding atmosphere is referred to as the motor with *optimum expansion ratio.*

Since changes in the atmospheric pressure affect the pressure thrust, a variation of the rocket-motor thrust with altitude is to be expected. The change in pressure thrust due to altitude changes amounts to 10 to 30 per cent of the overall thrust. Thus, the following conclusions may be stated:

1. External pressure decreases thrust.
2. Rockets would operate most efficiently

in a vacuum, where atmospheric pressure (P_a) is zero.

Now that we have covered the physics of jet-propulsion engines, let's see how these engines are classified into types.

CLASSIFICATION OF JET SYSTEMS

Your chief concern in regard to types of jet systems will be with jet-propulsion systems of two general classifications: rockets and ducted propulsion systems.

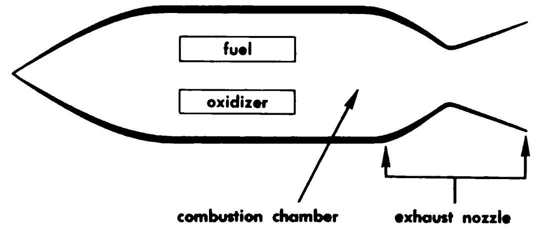
Rocket Systems

Rockets are self-contained; that is, all of the materials which are necessary for their operation are contained within the unit, as shown above right. The necessary materials are fuel and oxidizer. An oxidizer is a substance which contains the oxygen necessary for combustion of the fuel; therefore, the operation of a rocket unit is independent of the atmosphere. A rocket unit is sometimes spoken of as an air-independent system.

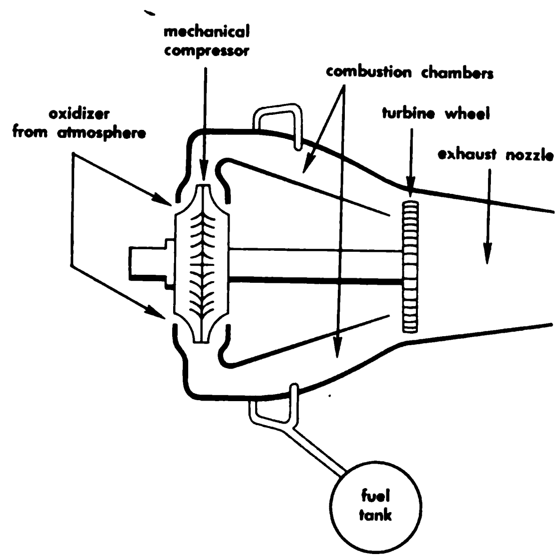
Ducted propulsion Systems

In the case of duct propulsion, the medium (air) through which the missile passes is taken into the unit and accelerated to a greater momentum by thermal means prior to ejection. One such type is the mechanical compressor unit in which the working fluid (air) is compressed, burned, and allowed to expand through an exhaust nozzle. The compression is usually achieved by some form of rotary compressor wheel, driven by a turbine. Study the sketches at right of the turbojet and turboprop engines as you read the next few lines. Hot gases that result from the combustion of fuel and air in the combustion chambers strike the turbine wheel causing it to rotate. The rotary motion is transferred by means of a shaft to the compressor wheel which compresses the air.

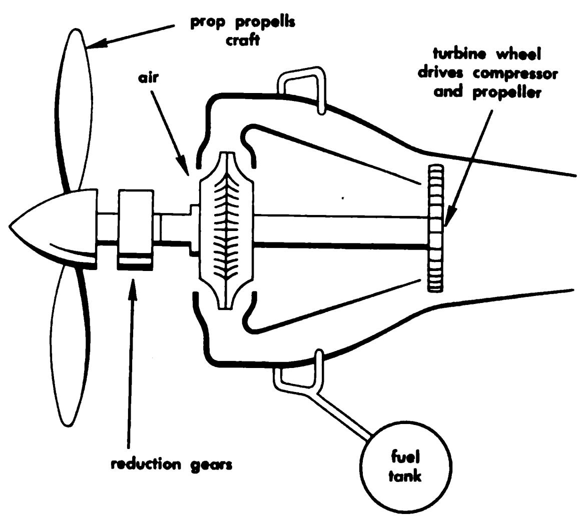
Another type of duct propulsion is the *pure duct*. In this type, compression of the working fluid is obtained by proper design of the intake section of the duct. Here again burning takes place and is followed by high-speed expansion through an exhuast nozzle. The ramjet engine pictured on the next page is this type.



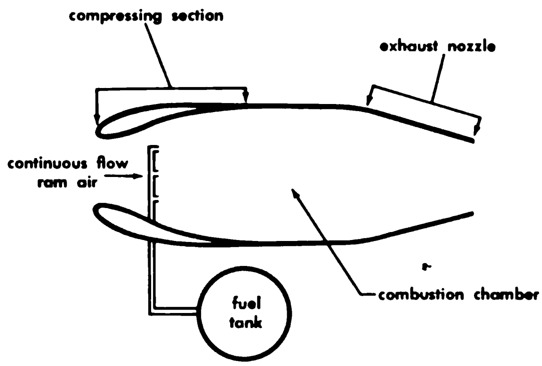
Rocket containing fuel and oxidizer (air independent)



Turbojet (air dependent)



Turboprop (air dependent)



Ramjet duct unit (air dependent)

A third type of duct propulsion is the intermittent, or pulsejet, unit. This engine also obtains pressure and velocity changes in the fluid medium by proper design of the intake section of the duct. It differs from the two previous types in that the flow of air and heat-giving fuel is *intermittent*, or in *cycles*. Notice the illustration below.

Now that you have the various types of jet-propulsion systems in mind, you are ready to consider the types of fuels these systems use.

PROPELLANTS USED IN GUIDED MISSILES

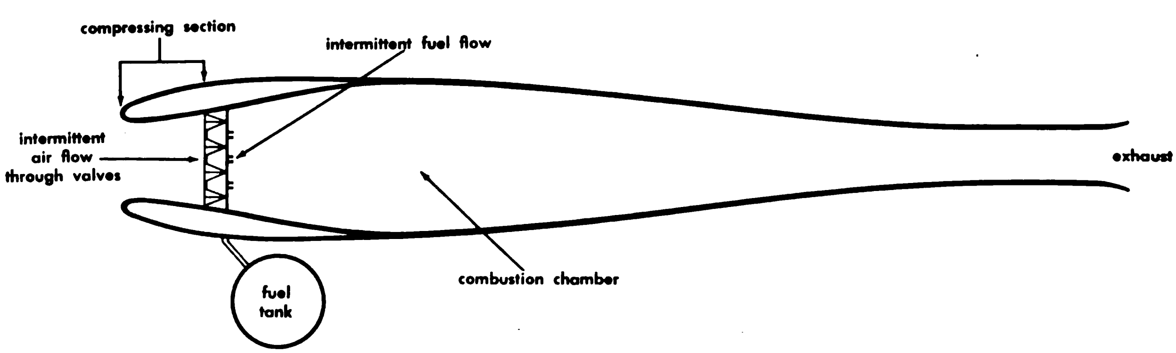
Earlier you read that thrust is developed as a result of energy released in the jet-propulsion engine. A large quantity of readily available energy is found in certain materials. When these materials react in a jet engine, they impart thrust to propel the missile. These materials and the oxidizers with which they react are called *propellants*. Large quantities of high-pressure and high-temperature gases are produced by the chemical reaction of a fuel and an oxidizer at the proper time and rate in the combustion chamber. The *heat energy* thus made available is converted into *kinetic energy* in the exhaust nozzle, or tail pipe.

When you read of an engine that can travel faster than an artillery shell, operate in a vacuum, deliver more energy than a reciprocating engine, and do so with few or no moving parts, you may get the idea that some very complex chemical mixture must be used as the propellant. Such is not the case. Jet-propulsion systems can operate on such common, cheap fuels as kerosene, gasoline, alcohol, gunpowder, and coal dust. However, obtaining a desired result depends upon, among other things, the proper choice of a *fuel-oxidizer* combination on the basis of energy-releasing capabilities and convenience of use.

From the standpoint of physical state, propellants may be either *solids*, *liquids*, *gases*, or various *combinations* of these. Generally the propellants are either solids, liquids, or a combination of the two. Gases are rarely used because greater energy transformation results when a substance goes from the solid or liquid state to the gaseous state than results when starting with gaseous propellants and accelerating them to new velocities.

Terms Referring to Propellants

From the standpoint of performance it is necessary to have a method of rating or com-



Intermittent duct (air dependent)

paring various propellants. Comparison is made by determining *total impulse.* Total impulse is the product of the thrust in pounds times firing duration in seconds. In formula we have:

I_T (Total Impulse in lb-sec) = T(Thrust in lbs) × t (Duration in secs).

Solid propellants are rated, or compared, on the basis of *specific impulse.* Specific impulse is the amount of impulse produced by one pound of the propellant. Stated in formula:

$$I_{sp} \text{ (Specific Impulse in lb-sec/lb)} = \frac{I_T \text{ (Total Impulse in lb-sec)}}{W \text{ (Weight of Solid in lbs)}}$$

A common method of comparing liquid propellants is on the basis of *specific thrust.* Specific thrust is equivalent to specific impulse for solid propellants but derived in a slightly different way. Specific thrust is defined as the thrust in pounds produced, assuming that the propellant is consumed at the rate of one pound per second. Specific thrust can be expressed:

$$T_{sp} \text{ (Specific Thrust in lbs/lb/sec)} = \frac{T \text{ (Thrust in lbs)}}{W \text{ (Weight Rate of Flow in lb. per sec)}}$$

Specific thrust is frequently expressed in seconds.

You sometimes may see the term *specific impulse* used for liquid as well as solid propellants. However, the term *specific thrust* more correctly brings out the correct physical meaning for liquids. As for solid propellants, it would be impractical to attempt to measure the weight rate of flow; therefore, specific impulse is used for comparison of solid fuels.

Specific propellant consumption is another term of importance in liquid-propellant systems. It is the reciprocal of specific thrust. It is defined as the propellant flow in pounds per second necessary to produce one pound of thrust; that is:

$$\text{Specific Propellant Consumption} = \frac{\text{Weight Rate of Flow}}{\text{Thrust}}$$

Other terms you should know are mixture ratio and exhaust velocity.

Mixture ratio designates the relative quantities of oxidizer and fuel used in the propellant combination. It is numerically equal to the weight of oxidizer flow divided by weight of fuel flow. *Exhaust velocity* is determined theoretically on the basis of the energy content of the propellant combination. The actual velocity of the exhaust gases is of course less than this theoretical value since no jet engine can completely convert the energy content of the propellant into exhaust velocity. Thus, *effective exhaust velocity* is sometimes used and is determined on the basis of thrust and propellant flow:

$$\text{Effective Exhaust Velocity} = \frac{\text{Thrust}}{\text{Mass Rate of Flow}}$$

Solid Propellants

In general, solid propellants consist of a fuel, usually a *hydrocarbon,* and an *oxidizer* which contains a large weight percentage of oxygen. These substances are mixed so as to produce a solid of desired chemical and physical characteristics. The finished product is called a *grain* or *stick.* One or more grains constitute a *charge.*

An ideal solid propellant would possess the following characteristics:

1. Manufactured from easily obtained substances.
2. Safe and easy to handle.
3. Easily stored: stable to shock and temperature changes.
4. Ignites and burns uniformly.
5. Maintains constant burning surface.
6. Nonhygroscopic (will not absorb water vapor).
7. Smokeless.
8. Flashless.

It is improbable that a single propellant having all these characteristics will be produced. Some of these characteristics are obtained at the expense of others, depending upon the performance desired.

CLASSIFICATION OF SOLID PROPELLANTS. Basically, solid propellant charges may be grouped under one of two types: restricted burning and unrestricted burning. A *restricted-burning* charge has some of its exposed surfaces covered with an inhibitor to control burning. By this procedure, burning can be

restricted to take place only on the desired surface or surfaces. Controlling the burning area in this manner lengthens burning duration and determines the combustion-chamber pressure for a given charge. A burning cigarette is a good example of a restricted-burning grain, if you consider the paper-covered portion as representing the inhibited area.

Unrestricted-burning charges are permitted to burn on all surfaces simultaneously. Relatively speaking, the unrestricted grain delivers a large amount of thrust for a short period of time, and the restricted-type grain yields a smaller amount of thrust for a longer period of time.

Burning of solid propellants may vary, depending upon the chemical composition, initial propellant temperature, combustion-chamber temperature, gas velocity adjacent to the burning surface, combustion-chamber pressure, and size and shape of the grain. One propellant grain may burn in such a manner that the burning area remains constant, producing constant thrust. This type of burning is known as *neutral burning*. Another grain may exhibit an increase in burning area as burning progresses. In this case, *progressive burning* is taking place. Still another grain may show an ever-decreasing burning area as burning progresses. This type is called *digressive burning*.

The *burning rate* of a solid propellant is the velocity at which the grain is consumed. The rate is a measure of the linear distance burned, in inches per second, in a direction perpendicular to a burning surface.

CONFIGURATIONS OF GRAINS. As stated earlier, thrust depends on mass rate of flow and change in velocity of the working fluid. For large thrust a large burning area is necessary in order to yield a large mass flow. A smaller burning area produces less mass flow and less thrust. Therefore, by varying the geometrical shape and arrangement of the charge, the thrust developed by a given amount of propellant in a given combustion chamber can be greatly influenced.

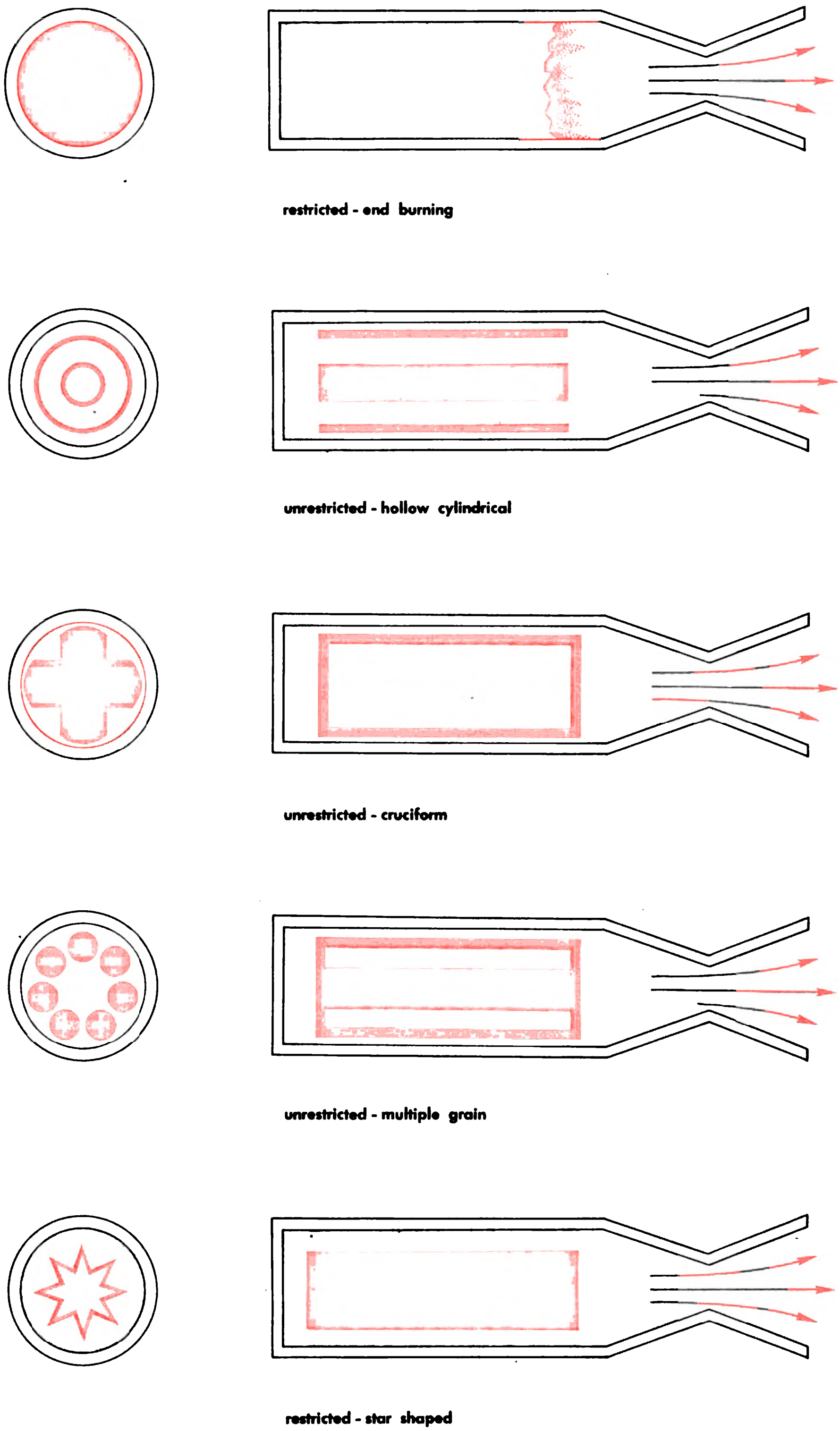
The drawings on the following page show some of the grain configurations in use. The top drawing shows a *restricted-burning* charge. It is a solid cylinder which completely fills the combustion chamber and burns only on the end face. Such a charge is called an *end burning* or *cigarette type*. The thrust is proportional to the cross-sectional area, and burning duration is proportional to the length.

The second drawing is an *unrestricted-burning* charge. Unrestricted charges are usually hollow and burn on both the outside and inside surfaces. Thrust is again porportional to the burning area. Since the inside area increases while the outside area decreases during burning, it is possible to maintain a nearly constant burning area. Burning duration, in the case of hollow grains, depends on the *web thickness*. Web thickness is the distance between the inside and outside surfaces. This is known as an *unrestricted hollow cylindrical charge*.

SOLID-PROPELLANT LIMITATIONS. Among the limitations of solid propellants is a grain's sensitivity to temperature. The initial temperature of a grain noticeably affects its performance. A given grain will produce more thrust on a hot day than on a cold day. One designed to produce 1000 pounds thrust at 80°F may deliver only 600 pounds thrust at 30°F. The initial temperature also has a definite effect on the burning rate. Such characteristics of grains emphasizes the need for storage under certain temperature conditions until time of usage. The percentage change of thrust per degree Fahrenheit temperature change is referred to as the *temperature sensitivity* for a particular propellant.

Temperature also affects the physical state of solid propellants. At extremely low temperatures, some solid-propellant grains become brittle and are subject to cracking. Cracks in a grain increase the burning area and therefore the combustion-chamber pressure. If this pressure exceeds the pressure for which the chamber was designed, a fracture or explosion results. A propellant exposed to high temperatures before firing may lose its shape, becoming soft and weak. This, too, results in unsatisfactory performance. The temperature range for most ordinary solid propellants is usually from about 25° F to 120° F.

Pressure limits play an important part in solid-propellant performance. Below a certain chamber pressure, combustion becomes highly



Solid propellant configurations

unstable. Some propellants will not sustain combustion at normal atmospheric pressure. Ordinarily, chamber pressure for solid propellants must be relatively high. For a given propellant composition and burning area, the chamber pressure is determined by the area of the exhaust-nozzle throat. If the throat area is too large, for example, proper chamber pressure cannot be maintained and an unstable condition occurs.

Decomposition and *hygroscopic* tendencies are weaknesses of solid propellants, but these weaknesses are minimized by the addition of other chemicals to the propellant base. Such chemicals are referred to as *additives*.

SPECIFIC SOLID PROPELLANTS. Some of the more common solid propellants are discussed below. Chemical formulas are listed in many cases in order that you may note the carbon and/or hydrogen content of the fuels and the high oxygen content of the oxidizers.

Black Powder. One of the first solid-propellant charges used was *black powder*. The approximate percentage composition follows:

Potassium Nitrate (KNO_3) (Saltpeter)	61.6%
Charcoal (C)	23.0%
Sulphur (S)	15.4%

Both charcoal and sulphur react readily with oxygen. Potassium nitrate, as you note from the formula, contains the oxygen. These substances are well mixed, using some substance such as glue or oil as a *binder*.

When heat is applied by a fuse of some sort, the potassium nitrate gives up oxygen. The oxygen reacts with the sulphur and carbon, producing intense heat and large volumes of carbon dioxide and sulphur dioxide gases. These gases constitute the major mass of the exhaust jet. The heat produced by the reaction gives high velocity to the exhaust gases. Black powder has a specific impulse of approximately 65 lb-sec/lb. It is not one of the more powerful propellants. One of its drawbacks is that it is quite sensitive to storage temperatures and tends to crack. Its exhaust velocity, relatively low, ranges from 1500 to 2500 ft/sec. It is used primarily in signal rockets and as an igniter for other solid-propellant grains.

Ballistite. Ballistite is a double-base propellant since it contains two chemical bases, *nitrocellulose* and *nitroglycerine*. It also contains small amounts of additives, each additive performing some specific function. A stabilizer absorbs the gaseous products of slow decomposition and reduces the tendency to absorb moisture during storage. A *plasticizer* serves as a binding agent. An *opacifier* is added to absorb the heat of reaction and prevent rapid thermal decomposition of the unburned portion of the grain. A *flash depresser* cools the exhaust gasses before the escape into the atmosphere, preventing a burning-tail effect. A typical ballistite composition follows:

Chemical	*Percent*	*Purpose*
Nitrocellulose $C_{24}H_{40}O_{20}(NO_3)$	51.38%	Propellant
Nitroglycerine $C_3H_5(NO_3)_3$	43.38%	Propellant
Diethylphthalate	3.09%	Plasticizer
Potassium Nitrate	1.45%	Flash Depresser
Diphenylamine	0.07%	Stabilizer
Nigrosine Dye	0.10%	Opacifier

Ballistite has a specific impulse of approximately 210 lb-sec/lb and is characterized by a smokeless exhaust. Since it decomposes over a period of time, storage temperatures between 40° and 120° F are desirable. The ingredients of ballistite are subject to detonation and are toxic when they come in contact with the skin. The manufacturing process is difficult and dangerous.

Galcit. A solid propellant produced by the Jet Propulsion Laboratory of the Guggenheim Aeronautical Laboratory of California Institute of Technology has been named JPL of GALCIT, the name resulting from using the first letters of the developing laboratory. *Galcit* by weight consists of about 25% asphalt-oil mixture which serves as fuel and binder and 75% potassium perchlorate ($KCLO_4$) which serves as an oxidizer. In its finished form, Galcit resembles stiff paving tar at ordinary temperatures.

Recommended temperature limits for firing are 40° F to 100° F. Its specific impulse is approximately 186 lb-sec/lb, and the manufacturing process is relatively simple. Storage-temperature limits are −9°F to +120°F. Thus, it is quite stable to temperature. It does not absorb moisture. A major disadvantage of Galcit is that the exhaust develops dense clouds of white smoke.

NDRC Propellants. NDRC propellants were developed through research by the National Defense Research Committee. A typical composition consists of approximately equal parts of ammonium picrate and sodium nitrate (46.5% each) and 7% resin binder (generally, urea formaldehyde). This propellant is noted for good thermal stability, but sodium nitrate absorbs moisture in atmosphere of high humidity, resulting in a soft and mechanically weak grain. It therefore must be sealed while stored. Heavy smoke formations develop in the exhaust gases.

Comparison of Solid Propellants. The table below gives you a general comparison of the last three solid propellants discussed above.

Liquid Propellants

In contrast to solid propellants which are stored in the combustion chambers of missiles, liquid propellants are *injected* into the combustion chamber from storage tanks. They permit *longer firing* duration and intermittent operation. Combustion can be stopped and started at desired intervals by controlling propellant flow.

Provided oxygen is used, or some oxygen-rich chemical, as an oxidizer, the best energy producing fuels are those liquids rich in *carbon* and *hydrogen*. Ethyl alcohol (C_2H_5OH) and aniline ($C_6H_5NH_2$) are examples.

In addition to fuel and oxidizer, a liquid propellant may also contain a *catalytic agent* to increase the speed of the reaction. *Inert additives* which *do not* take part in the chemical reaction but may produce higher thrust by increasing the propellant mass flow are sometimes used.

Many liquid-propellant combinations have been experimentally investigated, but, as in the case of solids, no combination has been found possessing all the desired characteristics. Desired characteristics are listed below:

a. Large availability of raw materials and ease of manufacture.

b. High heat of combustion per unit of propellant mixture to give a high chamber temperature.

	BALLISTITE	NDRC COMPOSITE	GALCIT
Specific Impulse lb-sec/lb.	210	160	186
Exhaust Velocity ft/sec	6800	5150	5900
Density lb/ft³	101.5	101	110
Temperature Sensitivity	High	Medium	Medium
Temperature Limits °F	Limited by high-temperature sensitivity	−40 to +140	−9 to +120
Burning Rate inches/sec	1.4	0.25	1.6
Chamber Temperature °F	5000 — 6000	3000 — 3500	3000 — 3500

c. Low molecular weight of the gaseous products of the reaction. Generally, propellants having a large weight percentage of hydrogen will best meet both "b" and "c" because of the much higher heat of combustion per pound of hydrogen than of carbon. Also, the hydrogen-containing products of the reaction have lower molecular weights than the carbon-containing products.

d. A low freezing point, permitting a wide range of operation.

e. High specific gravity, allowing a large weight of propellant to be stored in a given space and thus permitting smaller missile construction and less construction weight as a result of smaller, less-thick storage tanks.

f. Low toxicity and corrosiveness.

g. Low vapor pressure and no deterioration tendencies, thus simplifying storage problems.

CLASSIFICATION OF LIQUID PROPELLANTS. Liquid propellants may be grouped under either monopropellants or bipropellants. Monopropellants are those propellants in which the fuel and oxidizer are combined as one single substance. This combination is physical as in the case of hydrogen peroxide (H_2O_2) mixed with ethyl alcohol (C_2H_5OH), or it is chemical such as nitromethane (CH_3NO_2). Monopropellants are stable under ordinary temperature and pressure conditions. However, when activated by some initial ignition system, monopropellants decompose, liberating hot combustion gases. This action creates critical temperatures and pressures. Monopropellants are not frequently used because of their untrustworthy explosive characteristics. If they are disturbed, detonation could occur in fuel-delivery lines, possibly backing up into the missile storage tanks.

In the case of bipropellants, fuel and oxidizer are kept physically separated until they are injected into the combustion chamber. Combustion takes place under the usual high-temperature and-pressure conditions, producing high velocity exhaust gases. Most liquid propellant units are bipropellants.

A few of the many liquid fuels and oxidizers that have been investigated as possible missile propellants are discussed below in order to familiarize you with their properties and limitations.

LIQUID FUELS. Aniline, hydrazine hydrate, and ethyl alcohol are three of the more commonly used liquid fuels.

As used commercially, *aniline* ($C_6H_5NH_2$) is an oily clear liquid with a specific gravity of 1.022, making it slightly heavier than water. It has a boiling point of about 363°F and a freezing point of approximately 21°F. Since there is a large industrial demand for aniline for use in dyes, paints, and solvents, previously developed manufacturing processes make it easily obtainable. Aniline is manufactured by subsequent nitration and reduction of benzene. Upon contact with red fuming nitric acid, it ignites spontaneously. A fuel and oxidizer reacting in this manner are said to be *hypergolic*. This combination was successfully used in the WAC Corporal.

Aniline's relatively high freezing point can be lowered by the addition of up to 20% of furfural alcohol without altering the performance. It is noncorrosive, stable to shock, and burns similar to kerosene. It should be protected from sunlight to lessen deterioration. Aniline is a dangerous chemical, however, and requires extremely careful handling, since it can produce severe toxic effects in man through skin contact and by inhaling its vapors.

Hydrazine hydrate ($N_2H_4 \cdot H_2O$) is a colorless liquid, slightly heavier than water. It has a boiling point of 242°F, freezing point of −40°F, and an odor similar to ammonia. This substance fumes on contact with air and is explosive when its concentration is above 25 per cent. Hydrazine hydrate gives a hypergolic reaction with hydrogen peroxide and was used by the Germans as a component of the propellant for the ME 109 fighter plane.

Vapors from this fuel violently attack mucous membranes of the eyes, nose, and throat, and it irritates the skin. It attacks most common metals but is inactive to stainless steel and glass.

Ethyl alcohol (C_2H_5OH) is a clear liquid, lighter than water, and has a boiling point of about 173°F and freezing point of −178°F. It is stable to shock and temperature changes. This fuel is readily available because of its wide commercial market in the chemical and liquor industries. Alcohol is not ordinarily considered toxic but proves so if taken in-

ternally in excess. This fuel used along with liquid oxygen constituted the propellant of the German A-4 (V-2) rocket.

LIQUID OXIDIZERS. Liquid oxygen and forms of nitric acid are common oxidizers which are discussed below.

Liquid oxygen (O_2) is made by liquefying air and boiling off the nitrogen and other gases. This bluish-looking liquid has a boiling point of approximately −297°F and a freezing point of −363°F. It is heavier than water with a specific gravity of 1.14. Such a low boiling point causes an extremely high rate of evaporation. For this reason, storage and shipment to launching areas is quite a problem, resulting in appreciable loss. When poured on metal at ordinary temperatures it reacts like water dropped on a red-hot stove. Evaporation loss in the V-2 amounted to 4.4 pounds for every minute of time that elapsed between fueling and firing of the missile.

The extremely low freezing temperature of liquid oxygen causes water vapor from the surrounding atmosphere to collect and freeze on pipes and valves. Liquid oxygen is noncorrosive and nontoxic, but contact with the skin may produce severe burns. The term "burns" is used here in the same sense as you speak of burns resulting from dry ice.

Nitric acid (HNO_3) is used in several different forms as an oxidizer. One form, *mixed acid*, contains concentrated nitric acid (HNO_3) and a small amount of sulphuric acid (H_2SO_4).

White fuming nitric acid consists of concentrated nitric acid plus about 2 per cent of water.

Most frequently used and most powerful of the three is *red fuming nitric acid* (RFNA), consisting of concentrated nitric acid in which nitrogen dioxide (NO_2) is dissolved. It varies in color from orange to brick red and got its name from the reddish color of the nitric oxide fumes. It is easily obtained because of the large quantities commercially produced for explosives and fertilizers. RFNA is highly corrosive, so stainless steel is used for storage tanks and delivery pipes. Its high vapor pressure presents storage and transfer problems. The fumes are *extremely poisonous*, and *severe burns* result from bodily contact with the liquid. This oxidizer has been successfully used with aniline, giving up approximately 63.5% of its oxygen content for combustion.

MONOPROPELLANTS. A monopropellant of limited use is *nitromethane*. Nitromethane (CH_3NO_2) is a colorless liquid having a boiling point of approximately 214°F and a freezing point of about 19°F. Its use as a propellant is limited because of its tendency to propagate detonations through delivery pipes and storage tanks under certain conditions of temperature and pressure. As a monopropellant, it reacts best under relatively high chamber pressures. Initial combustion is usually set off by adding a small amount of gaseous oxygen and igniting by means of a spark plug. Nitromethane is not corrosive with most common metals and is only slightly toxic.

Hydrogen peroxide (H_2O_2) is a colorless liquid with freezing point of 14°F and boiling point of 288°F at 87% concentration. As sold at the corner drug store, it is a mild liquid of about 3 per cent concentration. As a monopropellant for jet-propulsion units, the concentration ranges from 70 to 90 per cent.

When in contact with some *catalysts*, such as sodium permanganate or potassium permanganate, it decomposes, forming steam and gaseous oxygen. When 90% hydrogen peroxide decomposes, about 42% of the total weight of the decomposition products is gaseous oxygen. Therefore, it is also used as an oxidizer with such fuels as alcohol and hydrazine hydrate.

A third use of this substance is as a pressurizing agent, in which case the gaseous products of decomposition are jetted against a turbine wheel which drives fuel and oxidizer pumps connected to the turbine shaft.

Containers for hydrogen peroxide are made of aluminum alloys, stainless steel, and certain plastics. It deteriorates in strength by only 1 to 3% when stored at moderate temperatures.

Splashes of concentrated hydrogen peroxide on the skin may produce chemical burns, depending on exposure time.

COMPARISON OF LIQUID PROPELLANTS. The following table on next page gives a general comparison of a few liquid propellants. The specific thrust and exhaust velocities are calculated values, assuming an exhaust gas expansion to one atmosphere of pressure.

Propellant (Oxidizer & Fuel)	Mixture Ratio O/F	Chamber Pressure, Psi	Chamber Temperature °F	Exhaust Velocity Ft/Sec	Specific Thrust, Sec
Liquid oxygen & liquid hydrogen	5.33	340	5,430	10,800	335
Liquid oxygen & hydrazine	0.50	300	4,500	8,350	259
Liquid oxygen & ammonia	1.4	300	4.950	8,220	255
Liquid oxygen & 100% ethyl alcohol	1.5	300	5,250	7,810	243
Liquid oxygen & gasoline	3.0	300	5,470	7,780	242
Liquid oxygen & 75% ethyl alcohol, 25% water	1.3	300	5,080	7,700	239
Gaseous oxygen & nitromethane	0.05	270	4,500	7,300	227
Red fuming nitric acid & aniline	3.0	300	5,020	7.090	221
Nitromethane	—	300	3.950	7,010	218
White fuming nitric acid & furfural alcohol	1.9	300	5.020	6,890	214
Hydrogen peroxide, 87% pure, plus water	—	300	1,310	4,060	126

Keep in mind that as yet no one propellant can be considered the best for missiles. The intended mission of a missile determines the kind of propellant used.

BASIC COMPONENTS OF JET-PROPULSION SYSTEMS

The basic components of any jet-propulsion system are the combustion chamber and the exhaust nozzle.

Combustion Chambers of Jet-Propulsion Systems

A combustion chamber is the enclosure in which the transformation of energy from potential to kinetic form occurs. Simple geometric shapes such as the cylinder and sphere are most common in the design and manufacture of the chamber. The length and diameter must be such as to produce a chamber volume most suitable for complete and stable combustion. The chamber length and nozzle *exit diameter* are determined by the propellants to be used. The chamber and nozzle exit must be designed to produce the proper gas velocity and pressure at the nozzle exit when a given propellant is used. Depending upon the type of propellant used, the combustion chamber may also contain an injection system and ignition system. Various types of these systems are discussed below.

INJECTORS. The function of an *injector* is similar to that of a carburetor in an internal-combustion engine. It *atomizes* and *mixes* the propellants in such a manner that a correctly proportioned fuel-oxidizer mixture results.

In the multiple-hole impingement injector illustrated on the page at right, oxidizer and fuel are injected through a varied arrangement of separate holes in such a way that the jet-like streams *impinge* (intersect each other at some predetermined point), breaking up into fine vapor-like droplets.

A spray injector, second illustration at the right, has oxidizer and fuel holes arranged in circles so as to produce *conical* or *cylindrical* spray patterns which intersect, thus becoming atomized and well mixed.

A nonimpinging type injector, lower illustration at the right, is one in which the oxidizer and fuel do not impinge at any specific point but mix mainly due to turbulence and the formation of propellant vapors. The V-2 uses such an injector head, in which case fine droplets of alcohol mix with gaseous oxygen.

IGNITION SYSTEMS. In order to initiate combustion, nonspontaneously ignitable propellants must be activated by absorbing energy beyond that which they already contain. This energy is supplied by an ignition system. An igniter must be located near the injector in such a manner that it receives a satisfactory starting mixture that readily ignites. Too great a collection of fuel and oxidizer in the combustion chamber before ignition may result in an uncontrolled explosion.

A *spark-plug ignition* system has been used successfully by locating the spark plug in a region where initial fuel and oxidizer vapors form an ignitable mixture. This type of ignition was used in the V-1 missile.

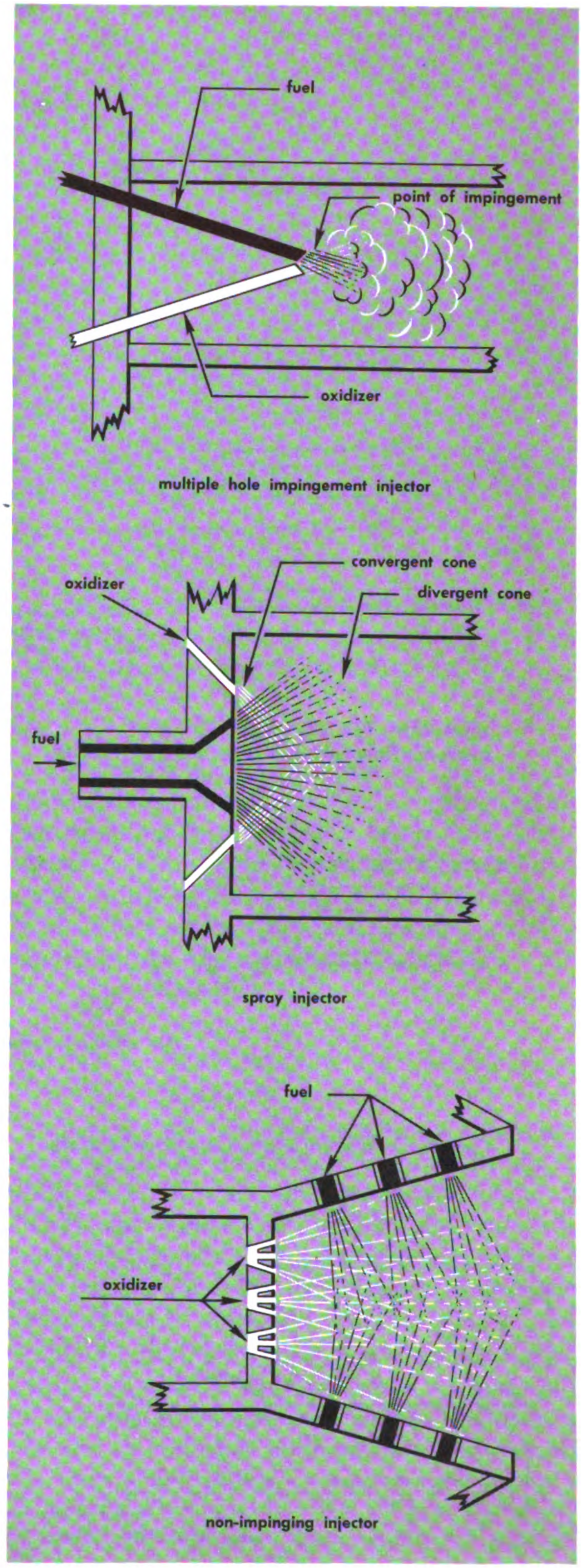
A *powder-charge ignition* system is used mainly for solid propellant charges. It consists of a powder squib which can be ignited electrically from a safe distance and then burned for a short time with a flame hot enough to ignite the main propellant charge.

A *catalytic ignition* method employs a solid or liquid catalytic agent which activates a chemical decomposition of the propellant, producing high-pressure exhaust gases.

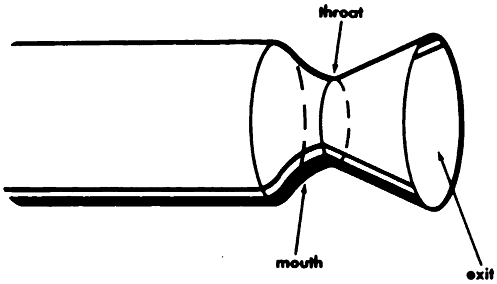
Exhaust Nozzles of Jet-Propulsion Systems

An exhaust nozzle is a nonuniform chamber through which the gases generated in the combustion chamber flow to the outside. In the nozzle, the most important areas to be considered are the cross-sections at the mouth, the throat, and the exit. These areas are indicated in the top illustration on the next page.

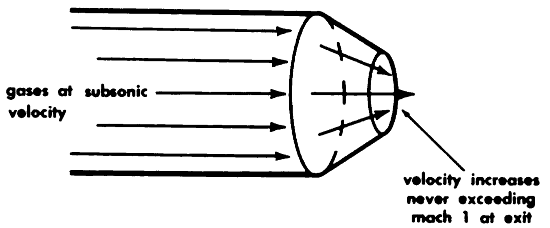
The function of a nozzle is to *increase* the velocity of the gases. Under conditions of steady flow, the weight of the gases which pass any cross-section in unit time is constant. (Bernoulli's theorem). Thus, in the case of



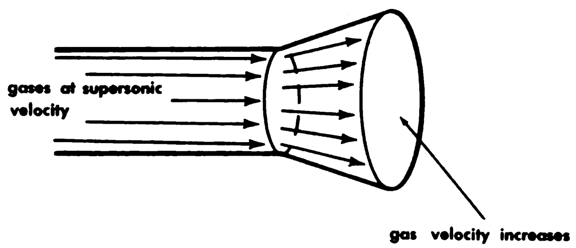
Types of injectors



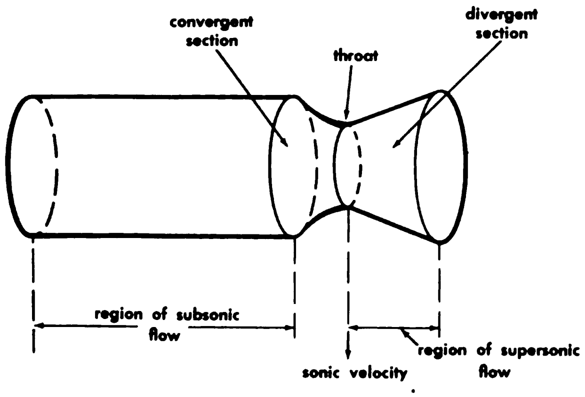
Location of nozzle components



Subsonic flow through a convergent nozzle



Supersonic flow through a divergent nozzle



Convergent-divergent, or "DeLavel," nozzle

subsonic flow, the velocity of the gases must increase if the cross-section is constricted at some point and the weight rate of flow stays constant. If the cross-section becomes wider, the velocity of the gases decreases. This relation of cross-section to velocity holds true for subsonic flow of gases but is not true for gases flowing at supersonic speeds.

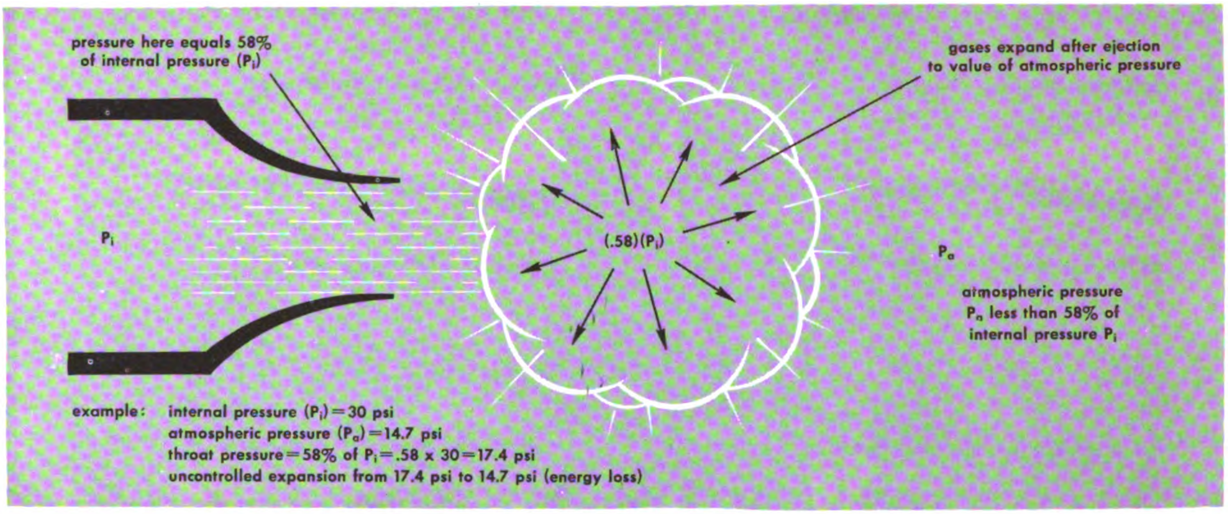
The velocity of subsonic gases passing through the convergent nozzle illustrated on the left would become higher and higher, ending with Mach 1 (local speed of sound) at the exit, assuming the nozzle were long enough. (Incidently, when reference is made to the local speed of sound in an exhaust nozzle, it does not mean the 760 mph value at sea level under standard conditions. Instead, it means the speed of sound that corresponds to the temperature at a specific point *in* the exhaust nozzle.)

Gases flowing at supersonic velocities, relative to the speed of sound at the mouth, are slowed down when passing through a convergent-type nozzle.

As shown in the sketch on the left, supersonic flow is increased when passing through a divergent nozzle. Subsonic flow, on the other hand, is decreased when passing through such a nozzle. This decrease of subsonic flow is in keeping with Bernoulli's theorem. Since the cross-section increases and the weight rate of flow remains constant, the velocity of flow must decrease with a proportionate increase in pressure. In the case of supersonic flow, however, the gases are in a compressed state. When the nozzle diverges it allows the gas to expand; the pent-up pressure energy is thus converted to kinetic energy, increasing the velocity of the gases.

To obtain a supersonic exhaust velocity, currently used rocket motors combine the convergent and divergent configurations as shown on the left. The exhaust nozzle first converges to bring the subsonic flow up to the local speed of sound, and then, at the right instant, the nozzle diverges, allowing the gases to expand. The expansion produces supersonic flow, as indicated in the illustration.

The convergent nozzle and the convergent-divergent nozzle, or DeLavel as the latter type is frequently called, are the two main



Uncontrolled expansion of gases outside nozzle

types of nozzles used currently in jet propulsion.

A convergent nozzle consists only of a convergent section with the throat located at the nozzle extremity. Tests on the flow of gases from this type of nozzle indicate that when 58 per cent of the internal pressure is a value greater than the atmospheric pressure, there remains an excess pressure in the gases as they emerge from the nozzle. The excess pressure and the uncontrolled expansion of gases outside the nozzle represent a loss of available energy. The above drawing illustrates this uncontrolled expansion of gases.

The convergent-type nozzle is, therefore, constructed for a specific set of propellant and combustion characteristics in order to attain the highest practical exhaust velocity.

A convergent-divergent nozzle can be used to control the expansion of the gases after they pass through the throat and thus obtain higher velocity and increased thrust. Again, the throat area is determined by the weight rate of flow desired. The area at the exit of the divergent section is determined by the ratio of expansion desired for the gases between the throat and the exit.

Other nozzle configurations which are of increasing importance are the *adjustable-area* types illustrated on next page. In these types, the nozzle area is varied to fit existing combustion conditions.

Having considered the basic components of jet-propulsion systems, we are ready to study some specific types of such systems.

AIR-JET ENGINES OF MISSILES

As mentioned in Chapter 1, any jet-propulsion system that obtains oxygen from the surrounding atmosphere to support fuel combustion is known as an air-jet engine. Pulsejets, ramjets, turbojets, and turboprops all fall in this category. These engines are often indicated by the following symbols.

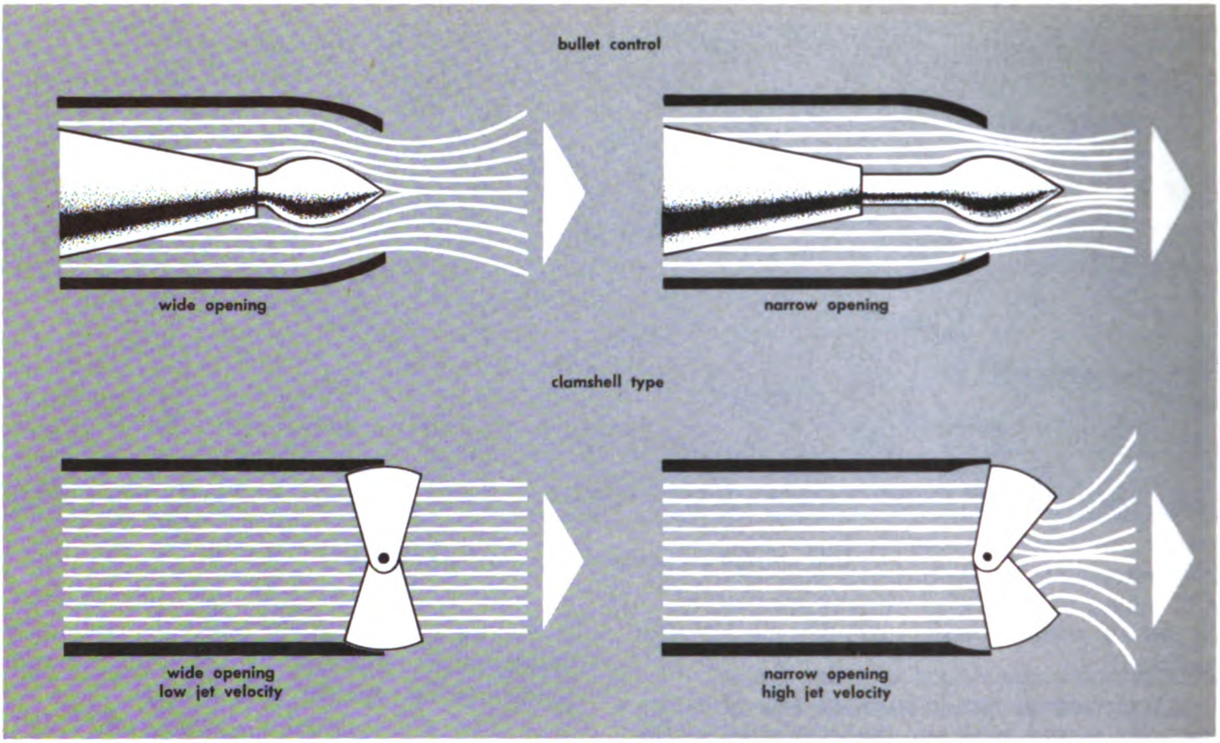
Pulsejet	PJ
Ramjet	RJ
Turbojet	TJ
Turboprop	T

You can see from the statement above how these engines' effective operation would be limited by the amount of oxygen available. They can operate only at altitudes where the oxygen content of the atmosphere is adequate for engine operation. The upper limit varies with different engines.

Before taking up the types of air-jet engines separately, look at the table on page 67. The table gives a general comparison of some of the operational characteristics of various jet-propulsion units.

Pulsejet Engine

Pulsejet engines received their name from the pulsating manner in which their com-



Adjustable area nozzles

bustion process takes place. This type of engine first drew international attention when used to propel the German V-1 missile, often referred to as the "Buzz Bomb."

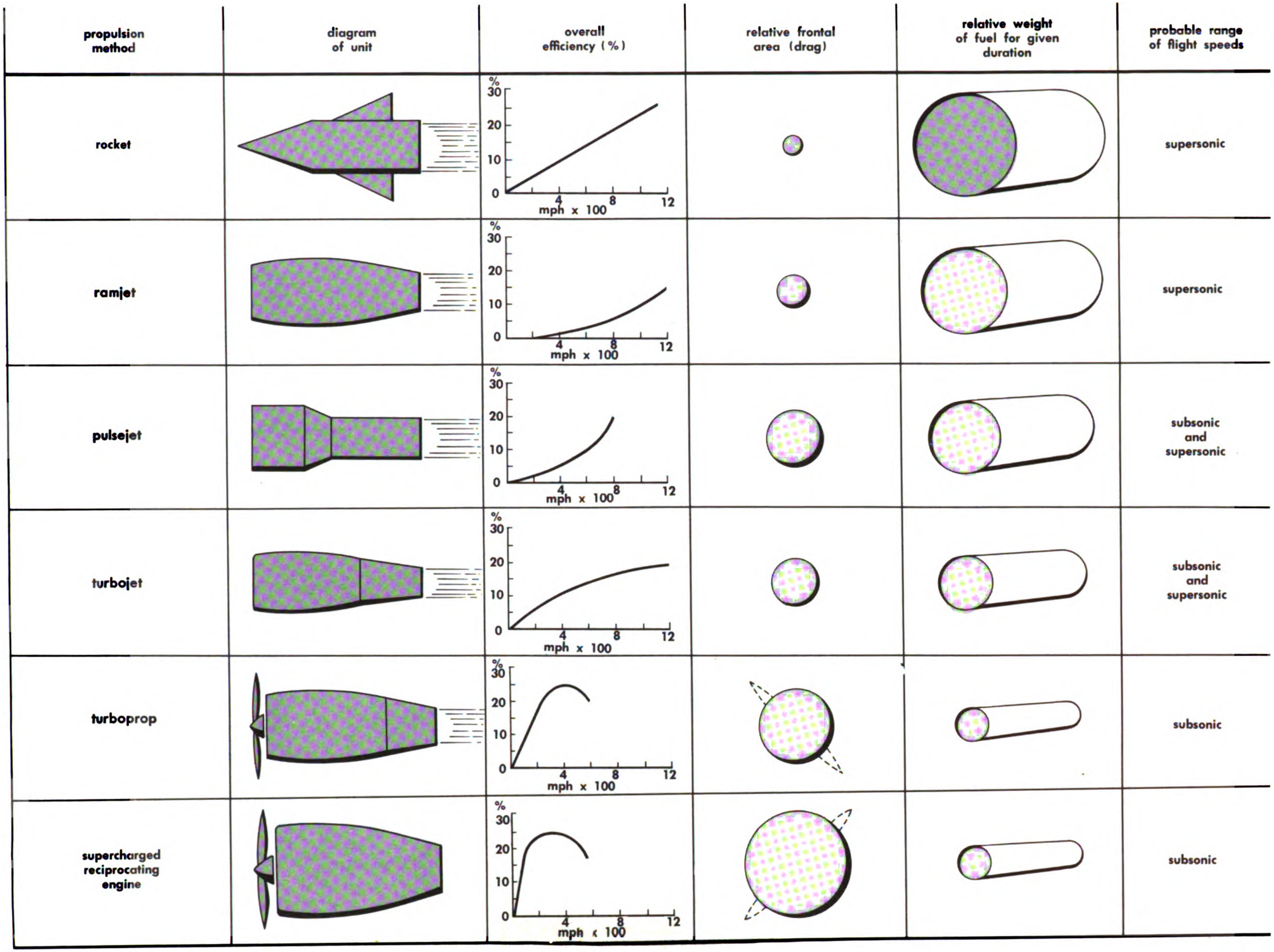
The American version of the V-1 is known as the JB-2. Research and development in the field of pulsejet propulsion since World War II has advanced the operational characteristics of such engines considerably, but the design remains basically the same. The pulsejet discussed below is the type used to power the JB-2. There are others of varied physical shapes; brief descriptions of some of the varied designs are given at the end of this section.

COMPONENTS OF A JB-2 ENGINE. The four major parts of a pulsejet engine are the diffuser, grill assembly (air valves, air injectors, fuel injectors), combustion chamber, and tail pipe.

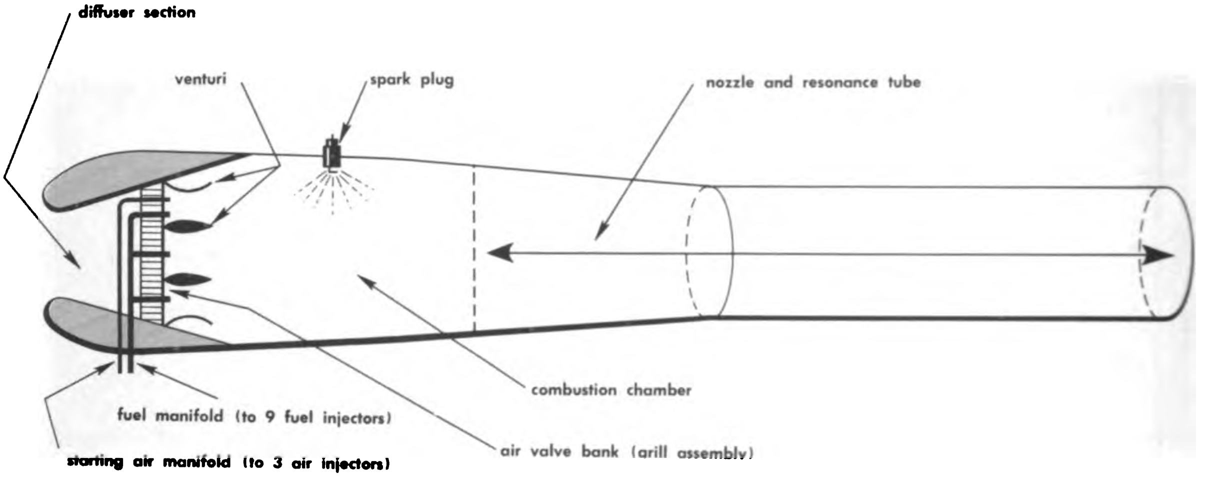
A *diffuser* is defined as a duct of varying cross-section, designed to convert a high-speed gas flow into low-speed flow at an increased pressure. The pulsejet air intake increases in cross-section from the mouth to the grill assembly. As air flows through this section, it undergoes a decrease in velocity and an increase in pressure. This increased pressure makes possible more rapid expansion when the air reaches the combustion chamber.

The *grill assembly* unit consists of a honeycomb arrangement of air intake valves, three starting air injectors, and nine fuel injectors.

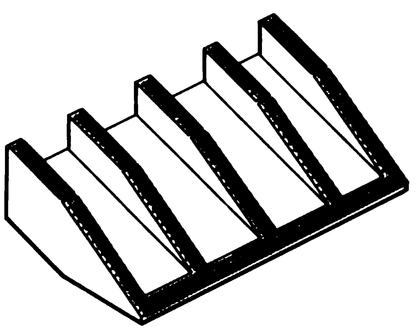
The design of air intake valves, or "flapper" valves as they are frequently called, is the most important feature of the grill assembly. These valves consist of V-section supporting members which are fitted with strips of spring steel. The open ends of the V's are toward the diffuser section and the closed ends point into the combustion chamber. The spring steel strips are spring-loaded in a *normally closed position*, that is, they exert pressure against their respective supporting members. When atmospheric pressure in the diffuser section is *greater* than combustion-chamber pressure, the flapper valves are forced open. When combustion-chamber pressure is *equal to*, or *greater than*, atmospheric ram pressure, the valves are closed. The two lower diagrams on page 68 show the construction and operation of one of the many air valves located in the grill assembly.



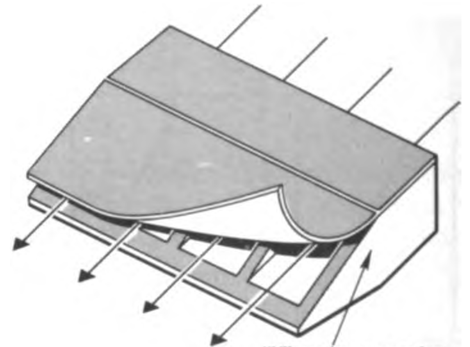
Operational characteristics of various jet-propulsion units



Structure of pulsejet engine

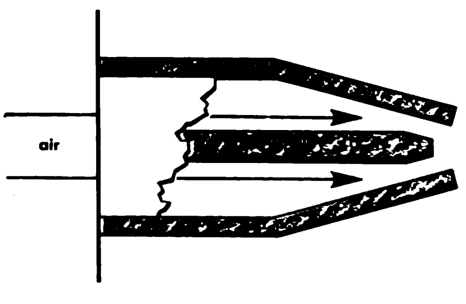


"V" section supporting member

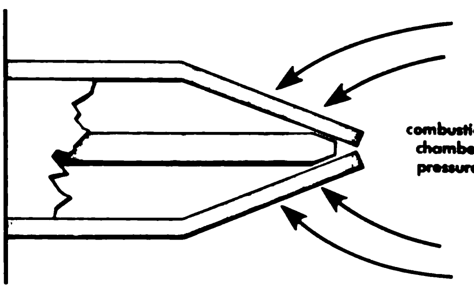


"V" section supporting member with spring steel leaf

Air valve construction

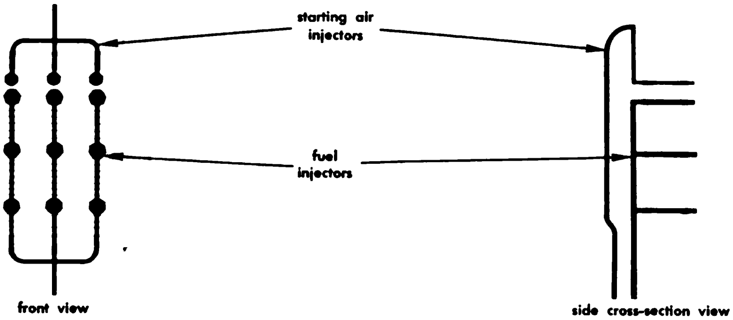


air valve open



air valve closed

Air valve positions



Grill assembly with air and fuel injectors

The flapper-valve unit is so designed that any number of layers can be assembled to form a grill of desired dimensions.

The three starting air injectors shown in the illustration above are connected to a pressurized external air supply which provides oxygen for *initial* combustion. They operate only during the starting phase — not during flight.

The nine fuel injectors also shown in the illustration above are fed through a delivery tube connected to the pressurized fuel supply stored within the missile.

Venturi sections, immediately aft of the grill assembly, insure proper fuel atomization and increased mixing of atomized fuel with air that enters through the flapper valves.

The *combustion chamber* contains the spark plug which provides initial ignition. Once started, combustion occurs periodically without spark-plug action.

The *tail pipe* of a pulsejet increases the velocity of the exhaust gases and determines the frequency of the combustion cycle. The frequency of operation is expressed by the following relationship:

$$\text{Frequency} = \frac{\text{Velocity of Sound}}{4 \times \text{Length of Tail Pipe}}$$

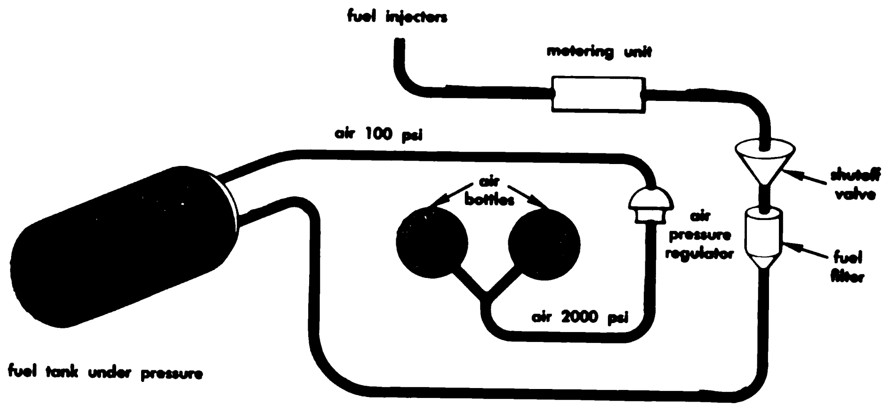
The frequency of operation of the JB-2 pulsejet is approximately 50 cycles per second. In small engines with short tail pipes, the combustion cycle frequency may easily exceed 200 cycles per second.

ENGINE AND FUEL-SYSTEM OPERATION. In a JB-2, fuel is carried in a steel tank of 180-gallon capacity. Fuel is pressure-fed to the fuel manifold. Notice in the first illustration on the next page that the two air bottles give an initial air pressure of 2000 pounds per square inch (psi) which is regulated to 100 psi by the air-pressure regulator. This air-pressure regulator keeps the air pressure on the fuel in the fuel tank from exceeding 100 psi. The fuel in the tank is forced through the fuel filter, the shutoff valve, and the metering unit to the fuel injectors.

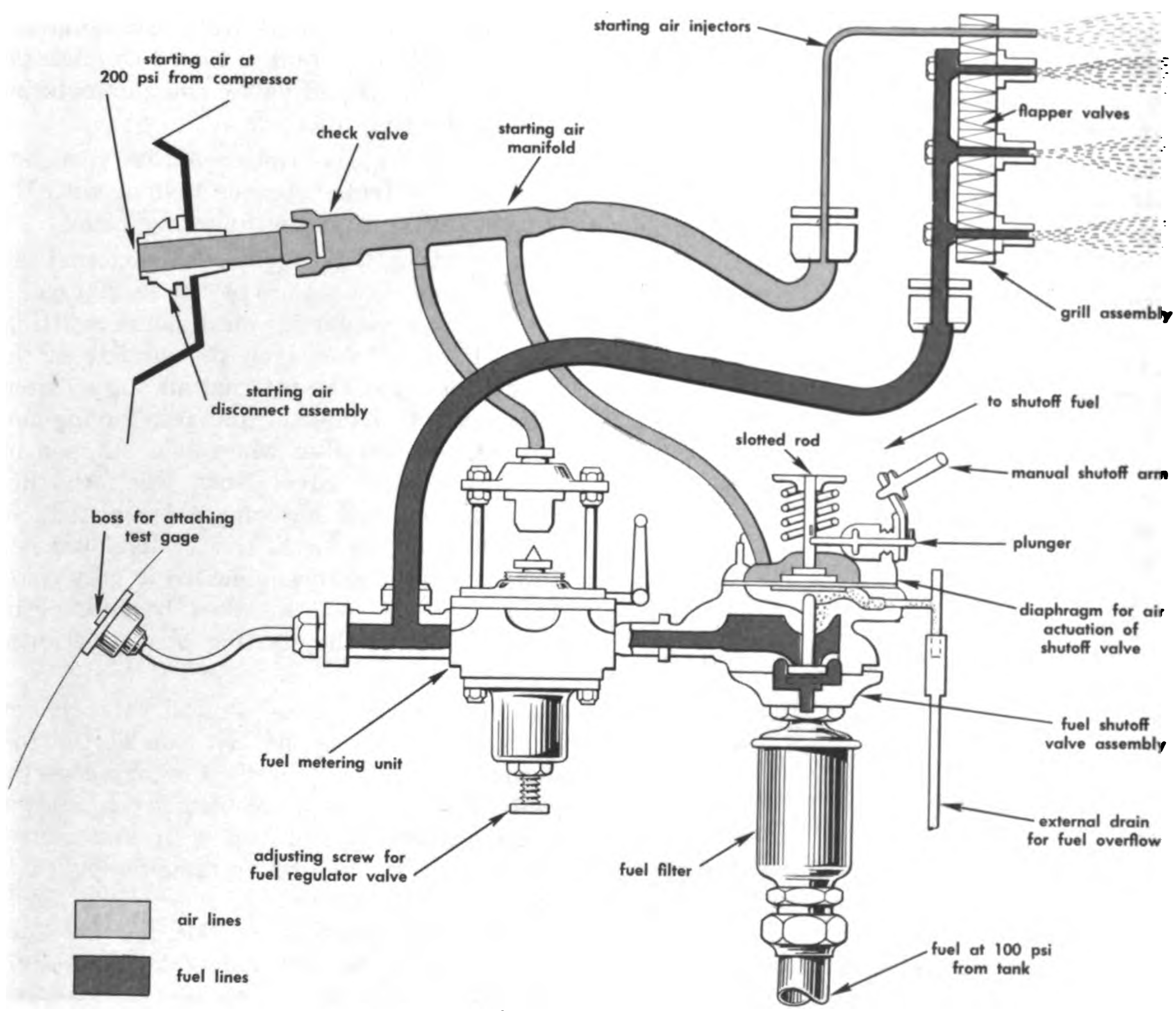
The following description of the operation of a pulsejet fuel system is tied in with the bottom diagram on the following page.

In starting this engine, an external air supply, under a pressure of 180 to 200 psi, is connected to the starting air disconnect fitting. This fitting is located on the outside of the missile fuselage. The internal air supply pressurizes the fuel tank at 100 psi, forcing fuel through the fuel filter where it is stopped by the fuel shutoff valve. Now, when the line from the external air supply is opened, air under pressure forces the check valve open and flows through the lines indicated in grey color. This "starting air," so-called because of its use only during the starting phase, performs three functions:

1. Air enters the fuel shutoff valve assembly and pressurizes the top side of the fuel shutoff diaphragm. This diaphragm pushes the piston-type fuel shutoff valve *down* to a lower position, allowing fuel to flow up and around the valve. This fuel valve remains locked in the *open* position by the plunger connected to the manual shutoff arm. The plunger slips into a slot as the rod above the diaphragm goes down with the diaphragm and shutoff valve. In case of malfunction during starting, the fuel may be cut off by pushing the manual



Fuel and pressure system of pulsejet engine



Pulsejet fuel system

shutoff arm which pulls the plunger from the slotted rod, allowing the diaphragm and fuel valve to rise.

2. Air enters the cylinder above the fuel metering unit where it extends a plunger downward to contact the fuel metering arm inside (not visible in diagram), thus restricting the fuel flow to the grill to a predetermined amount. This operation prevents flooding of the engine during starting.

3. Air flows through the three starting air injectors (previously illustrated on page 69) into the combustion chamber, where it mixes with the fuel to form a combustible mixture.

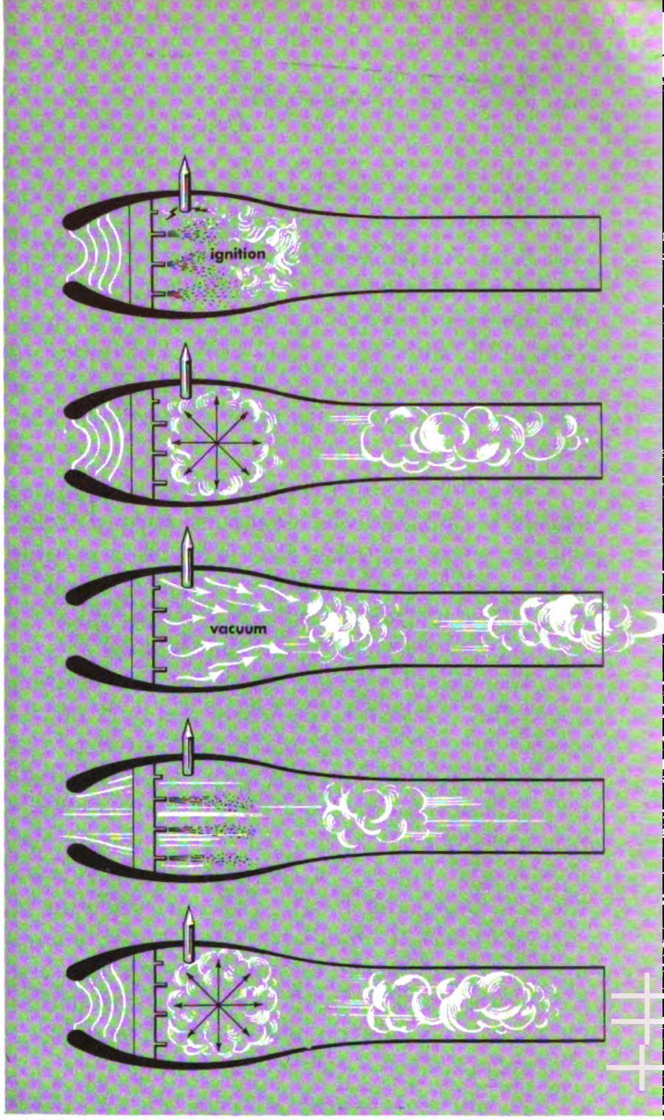
The spark plug is connected to a 10,000-volt ignition voltage source, and the fuselage is electrically grounded.

The starting air valve is opened and the ignition switch closed at approximately the same time. As stated above, the starting air allows fuel to pass through the fuel shutoff valve, metering unit, and nine fuel injectors where the fuel mixes with air from the three air injectors. The fuel regulator adjusting screw is preset to control fuel flow to the fuel injectors. The fuel-air mixture is ignited by the spark plug. Spark plug and starting air supply are no longer needed once combustion is started.

When the initial charge of fuel and air is ignited, pressure forces, created by the rapid combustion, accelerate the gaseous products in all directions. Since the flapper valves are closed, the high-velocity gases going in that direction cannot escape. They do, however, exert a great force against the grill assembly. This force is a component of thrust, and it also acts on the nine fuel-injector openings, preventing fuel from flowing during the high-pressure stage of the combustion reaction.

The diagrams above right indicate the various stages in the operation cycle of the JB-2-type pulse-jet engine.

In the first diagram, the flapper valves are closed. Pressure on both sides of the grill is the same. Starting air from externally pressurized source enters through the three starting air injectors; fuel under pressure enters through the fuel injectors; and the spark plug ignites the air-fuel mixture.



Stages of pulsejet engine

The fuel-air mixture burns rapidly; combustion-chamber pressure rises to prevent fuel flow and hold flapper valves in closed position; gases begin to expand.

Flaming gases rush down the tailpipe, and a partial vacuum (low-pressure area) is created in the combustion chamber.

The flapper valves open as shown in the fourth diagram, admitting a new charge of air (combustion-chamber pressure lower than diffuser pressure); a new charge of fuel enters through fuel injectors (combustion-chamber pressure lower than fuel pressure). Part of the exhaust gases return to the combustion chamber, igniting the fresh charge.

The fuel-air mixture burns rapidly; combustion-chamber pressure rises; the flapper valves close; fuel flow stops; and the gases begin to expand.

Since there is no resistance toward the rear, the rapidly expanding gases rush down the tail pipe. Actually, they overexpand, creating a partial vacuum in the combustion chamber. At this point in the cycle, atmospheric pressure in the diffuser section, or ram air pressure as it is called when the engine is moving, is again greater than combustion-chamber pressure.

The difference in pressure forces the flapper valves to open, admitting a fresh supply of air. At the same time, fuel-line pressure is greater than combustion-chamber pressure. Thus, a new charge of fuel enters through the injectors. The partial vacuum in the combustion chamber also has a suction effect on the flaming gases rushing down the tail pipe. Part of the gases reverse their direction, flow back to the combustion chamber, and ignite the fresh fuel-air charge. Again the combustion-chamber pressure rises. The flapper valves close; fuel flow stops. The hot gases expand, rush down the tail pipe, and create a partial vacuum in the combustion chamber. The flapper valves open, fuel flow starts, and flaming exhaust gases are sucked back to the combustion chamber to ignite the new charge. *The above is the cycle of the pulsejet engine.*

The pulsejet engine as used on the JB-2 does not develop enough static thrust for take-off under its own power. It must be boosted to operating speed by some type of launching mechanism.

The velocity of this engine is limited by the operation of the air valves. If the speed becomes too great, ram air pressure in the diffuser section is always greater than combustion-chamber pressure; consequently, the valves cannot close.

This engine is cheaply and simply constructed, economically operated, and applicable as a power plant for target drones and test vehicles.

Shrouded Pulsejet Engine

Successful operation of conventional pulsejets has been limited so far to comparatively low flight velocities. As the velocity is increased, the increasing ram pressure at the inlet valves requires a higher combustion pressure to keep the valves closed over a sufficiently long fraction of each cycle; at the same time, the back flow through the tail exit and precompression in the combustion chamber are reduced. It becomes more and more difficult for the combustion process to maintain the pressures observed during static operation of the engine, let alone maintain pressures high enough to equalize the ram air pressure on the diffuser section. Eventually, the valves remain open for too great a fraction of each cycle. At about a flight Mach number of 0.6, the engine ceases to operate in its pulsating manner.

Attempts are being made to extend the useful operation range of pulse-jets to higher flight Mach numbers. One such approach is to place the pulsejet engine inside a *shroud*, which is designed to keep the flow around the engine at a low Mach number at all times. In a power plant of this type, called a *shrouded* or *ducted* pulsejet, the troublesome pressure differences between *ram pressure at the inlet valves* and *static pressure at the tail-pipe exit* are practically eliminated. And with a suitable shroud design, it should be possible to keep the engine operating at any flight Mach number. Furthermore, this scheme allows the primary engine to make full use of the possible ram precompression which is not utilized in conventional pulsejets.

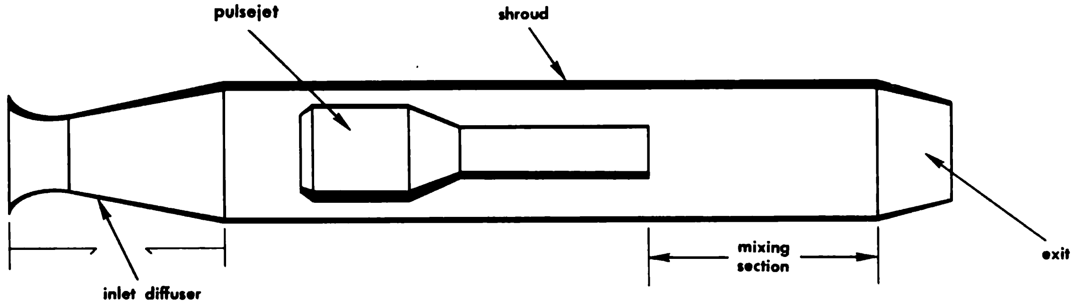
Construction of a shrouded pulsejet is shown in the illustration at top right.

Air enters the shroud through the inlet diffuser where it is slowed down to a low Mach number. Part of this flow is required for the operation of the pulsejet. The exhaust from the pulsejet tail pipe is mixed with the remaining shroud flow. Finally, the gases are returned to the atmosphere through the exit nozzle of the shroud.

Ramjet Engine

A ramjet engine derives its name from the ram action which makes possible its operation. This engine is sometimes referred to as the *athodyd* meaning aerothermodynamic duct.

Theoretically, ramjet operation is limited to altitudes below 90,000 feet because atmospheric oxygen is necessary for combustion. The velocity that can be attained by a ramjet engine is unlimited, theoretically. Actually,



Basic structure of a shrouded pulsejet engine

the faster it travels the better it operates and the more thrust it develops. Its speed is limited, however, at approximately Mach 5.0 because the skin temperature as a result of frictional heating has a harmful effect on the metals used in construction. The major disadvantage of a ramjet power plant is that the higher the operational speed for which it is designed, the higher the speed to which it must be boosted before initial starting of the engine. The speed range in which ramjets are designed to operate is the basis for their classification; that is, they are classified as subsonic or supersonic ramjets.

COMPONENTS OF A RAMJET ENGINE. Basically, a ramjet consists of a cylindrical shaped tube open at both ends, with a fuel-injection system inside. The engine is extremely simple in design and has no moving parts. Even though all ramjets contain the same basic components, the structure of these components must be modified to produce satisfactory operation in various speed ranges. The necessary modifications are explained later in the discussions of the subsonic, low-supersonic, and high-supersonic ramjets.

The component parts of a ramjet engine are:

Diffuser section

Combustion chamber with the following parts:

Fuel injectors

Spark plugs

Flame holder

Exhaust nozzle

The *diffuser section* serves the same purpose in the ramjet as it does in the pulsejet. It decreases velocity and increases the pressure of the incoming air. Since there is no wall or closed grill in the front section of this engine, the pressure increase of the ram air must be great enough to prevent the escape of the combustion gases out of the front of the engine. Diffusers must be especially designed for a given entrance velocity, or predetermined ramjet speed. In other words, the desired pressure barrier is developed only when air is entering the diffuser at the speed for which that particular diffuser was designed. The ideal situation would be a ramjet using a diffuser design which could be automatically altered in flight to conform with any ramair entrance velocity.

The *combustion chamber* is of course the area in which burning occurs and high-pressure gases are formed. Contrary to the pulsejet, the ramjet utilizes a continuous flow of fuel and air, and combustion is continuous. The fuel injectors are connected to a continuous-flow fuel-supply system, adequately pressurized to permit fuel flow in the midst of high-pressure gases which exist in the forward section of the combustion chamber. Combustion is initially started by spark-plug ignition, but once started it is continuous and self-supporting. The flame holder prevents the flame front from being swept too far toward the rear of the engine, thus stabilizing and restricting the actual burning to a limited area. The flame holder also insures that the combustion-chamber temperature will remain high enough to support combustion.

The *exhaust nozzle* performs the same function as in the case of the pulse-jet or any other jet-propulsion engine.

STRUCTURE AND BASIC OPERATION OF SUBSONIC RAMJETS. A subsonic ramjet engine

cannot develop static thrust; therefore, it *cannot* take off under its own power. If fired at rest, high-pressure combustion gases would escape out the front as well as out the rear. For satisfactory operation, the engine must be boosted to a suitable subsonic speed so that the ram air entering the diffuser section develops a pressure barrier high enough to restrict the escape of combustion gases to the rear only. Note the simple tubular construction and the openings at the front and rear in the illustration at upper right of a subsonic ramjet.

As ram air passes through the diffuser section as indicated in the preceding diagram, the velocity of the air decreases while the pressure increases. This is brought about by the increase in cross-section of the diffuser (Bernoulli's theorem for incompressible flow). Fuel is sprayed into the combustion chamber through the fuel injectors. The atomized fuel mixes thoroughly with the incoming air, and the mixture is ignited by the spark plug. As previously stated, after initial ignition, burning is continuous and no additional spark-plug action is necessary.

The gases which result from the combustion process expand in all directions as shown by the arrows in the central portion of the combustion chamber. The gases, expanding in the forward direction, are stopped by the barrier of high-pressure air and the internal sloping sides of the diffuser section, as indicated in the diagram by the short, wide, black arrows. The only avenue of escape remaining for the combustion gases is through the exhaust nozzle, and here another important energy conversion occurs. That is, the pressure energy of the combustion gases is converted to velocity. To be more specific, the high-pressure combustion gases enter the exhaust nozzle with a velocity which is below the local speed of sound. However, while passing through the convergent-type nozzle, the pressure energy of the gases decreases and the velocity increases up to the local speed of sound at the exhaust nozzle exit.

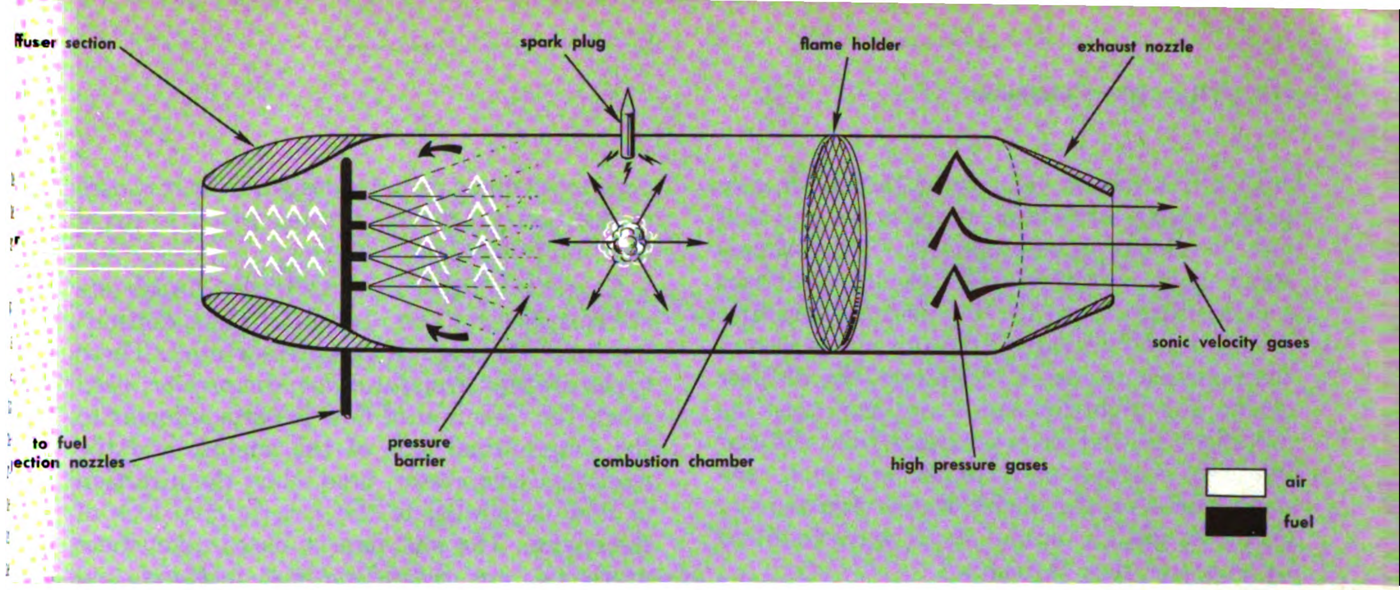
Thrust is developed in the ramjet as a result of the unbalance of forces acting in forward and rearward directions. The bombardment of combustion gases against the sloping sides of the diffuser and the ramair barrier exert a force in the forward direction. This forward force is not balanced by the combustion gases which escape through the exhaust nozzle. The unbalanced force determines the thrust.

STRUCTURE AND BASIC OPERATION OF LOW-SUPERSONIC RAMJETS. In order to operate, a low-supersonic ramjet must be boosted to a supersonic speed, approximately equal to operating speed, before ignition. When the forward speed of the ramjet becomes supersonic, a normal shock wave forms at the entrance to the diffuser section. The location of the normal shock wave is indicated in the diagram at center right. On the up-stream side of this normal shock wave the free-stream air is moving at a low-supersonic velocity. As the supersonic air passes through the shock wave, its velocity drops abruptly to a subsonic value with a corresponding increase in pressure. Thus, the shock-wave formation produces a valuable increase in air pressure at the diffuser entrance. As the compressed subsonic air flows through the diverging-type diffuser section, an additional increase in pressure and decrease in velocity is produced. The highly compressed air is now ready for combustion.

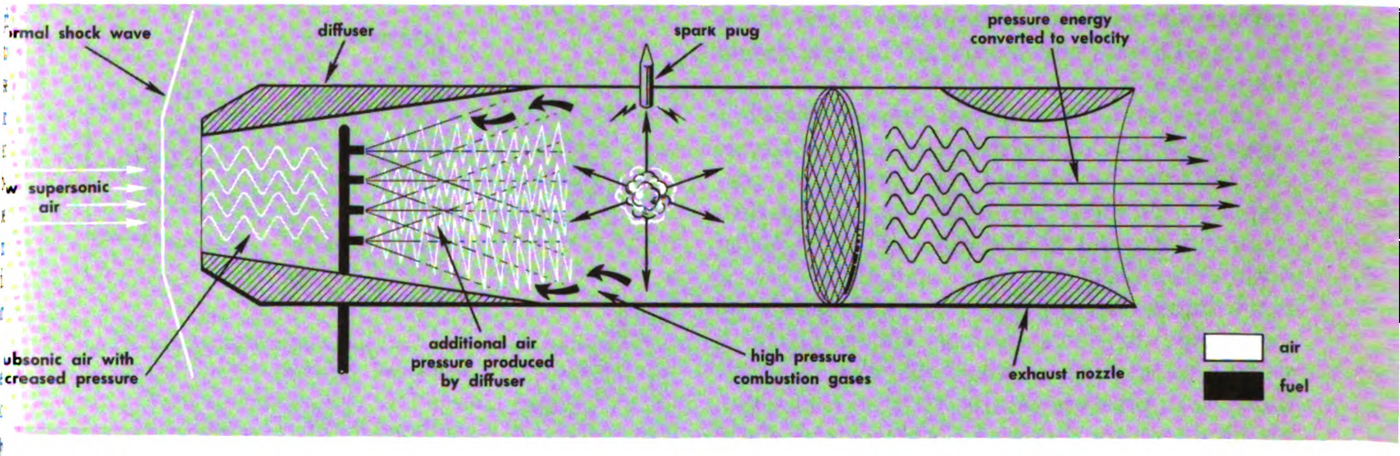
The combustion process is essentially the same as the combustion process in a subsonic ramjet. Fuel is mixed with the highly compressed air, the mixture is ignited initially by means of a spark plug, and burning is continuous thereafter. The potential energy possessed by the combustion gases is converted into kinetic energy by the exhaust nozzle.

The convergent-divergent nozzle shown in the diagram allows the gases to exceed the local speed of sound. Therefore, with proper design modifications, the ramjet engine can travel efficiently at supersonic speeds.

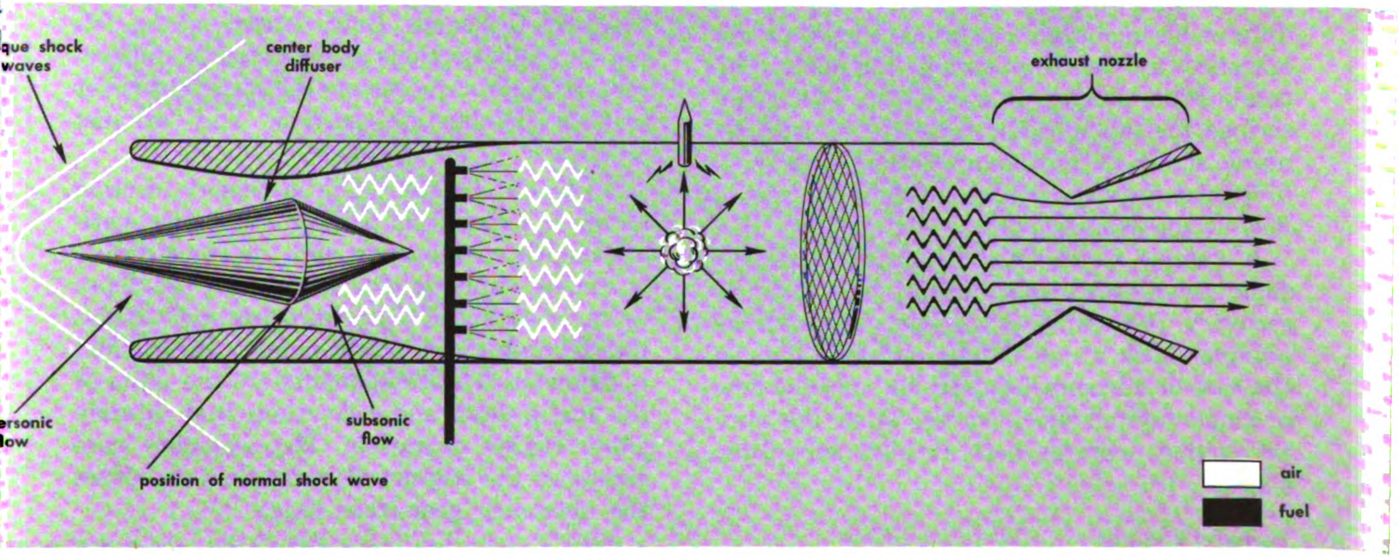
STRUCTURE AND BASIC OPERATION OF HIGH-SUPERSONIC RAMJETS. Now, suppose we desire a ramjet that will travel at much higher supersonic speeds, say Mach 2.0. At speeds in the neighborhood of Mach 2.0, shock waves formed at the diffuser inlet would be oblique and not normal. Air velocity in front of an oblique shock wave is high supersonic. When supersonic free-stream air passes through an oblique shock wave, an increase in pressure and a decrease in velocity occur, but the velocity is still supersonic. For example, air



Structure and combustion processes of a subsonic ramjet



Structure and combustion processes of a low supersonic ramjet



Structure and combustion processes of a high supersonic ramjet

with a free-stream velocity of 1500 mph may pass through an oblique shock wave and still have a velocity of 900 mph. Also, when supersonic air flows through divergent-type diffuser sections as shown in the first two ramjet diagrams, the velocity of that air increases and the pressure decreases. Therefore, the diffuser design for high-supersonic ramjets must be modified, so that in progressing from diffuser inlet to combustion-chamber entrance, the obliqueness of the shock waves successively decreases until a normal shock wave followed by subsonic flow is realized.

This energy transformation is achieved by utilizing a diffuser design of the type at the bottom of the preceding page. The center body diffuser decreases the obliqueness of the shock waves, allowing supersonic air to flow inside the diffuser inlet.

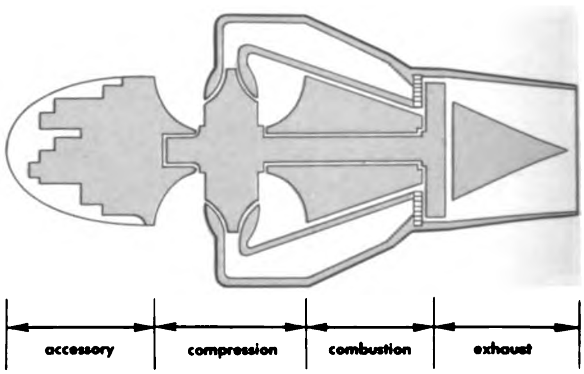
As supersonic flow passes through the convergent section of the diffuser, the velocity is steadily decreased and the pressure correspondingly increased. However, at some predetermined point in the diffuser, air velocity approaches the sonic value and a normal shock wave forms. As previously stated, when low-supersonic air flows through a normal shock wave, an abrupt decrease in velocity and increase in pressure results. Now, the subsonic air produced by the normal shock wave flows through the divergent section of the diffuser where it undergoes an additional velocity decrease and pressure increase. Here again the diffuser has achieved a pressure barrier at the entrance to the combustion chamber. The combustion process is the same as that described for the subsonic ramjet. The exhaust nozzle shown in the diagram is the convergent-divergent type designed to produce supersonic flow at the exhaust-nozzle exit.

A ramjet is designed to operate best at some given speed and altitude. The pressure recovery process in a diffuser designed for oblique shock waves is more efficient than in the case of diffusers designed for subsonic flow or single normal shock waves. Thus, the ramjet engine operates best at high supersonic speeds.

Turbojet Engines

A turbojet engine is an air-dependent thermal jet-propulsion device. The turbojet derives its name from its design, in that it uses an exhaust-gas-driven turbine wheel to drive its compressor.

Classification of turbojets. Turbo-jets are classified into two general groups, depending upon the type of compressor used. These types are the *centrifugal-flow compressor*, used when the direction of flow is perpendicular to the longitudinal axis of the engine, and the *axial-flow compressor*, used when the direction of flow is parallel to the longitudinal axis of the engine. The jets' names, based upon these processes, are *centrifugal-flow turbojets* and *axial-flow turbojets*. The illustrations below and top right show each type of turbojet engine.

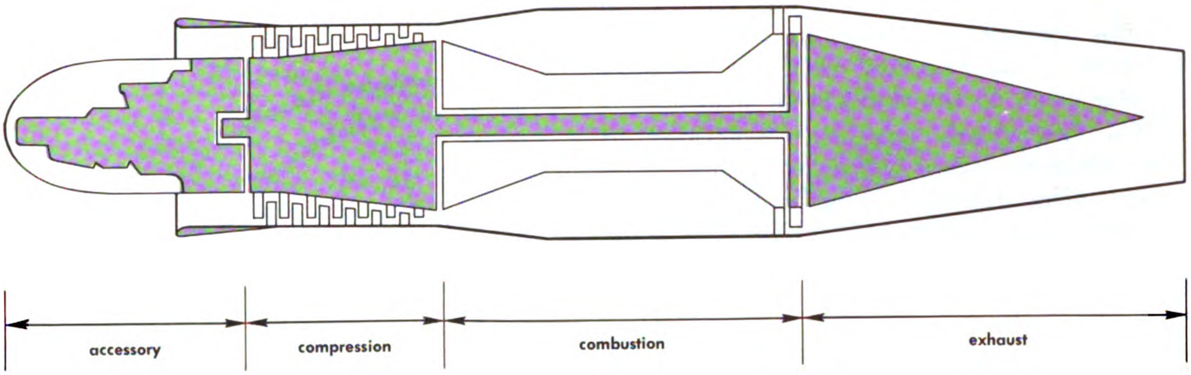


Centrifugal flow turbojet

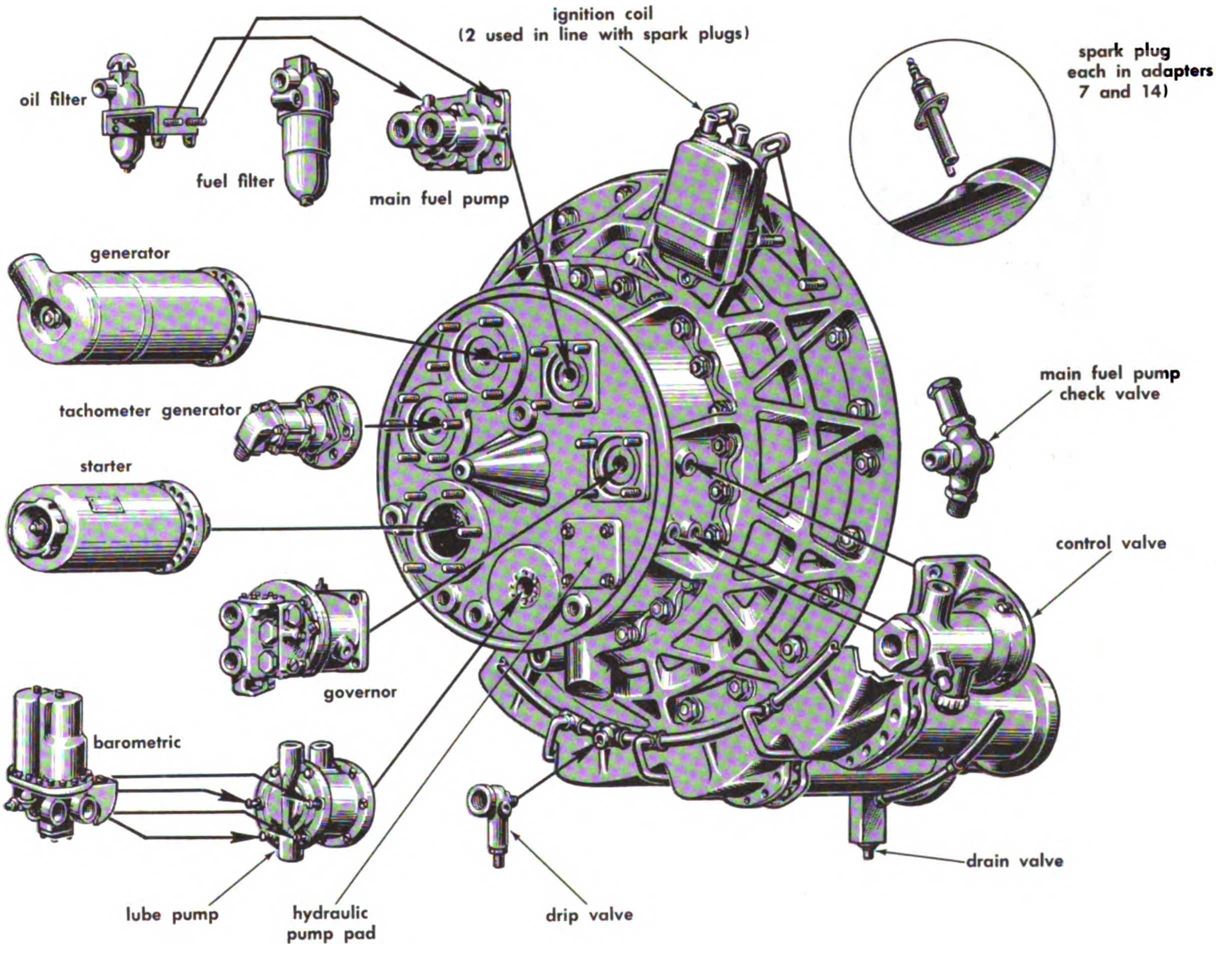
Component parts of turbojets. The operation of both types of turbojet engines is basically the same. Both consist of the following major sections as indicated in the preceding diagrams; accessory section, compressor section, combustion section, and exhaust section.

Accessory Section. The accessory section serves as a mounting pad for accessories, various engine components, and the front engine balancing support. The accessories are those units not essential to engine operation, such as the generator, hydraulic pump, starter, and tachometer. The components are the units of the fuel system and oil system that directly affect engine operation. The accessory case serves as the engine oil reservoir and houses the accessory gear-train cage.

The accessories comprising the accessory section of a typical turbojet engine are shown in the diagram at right.



Axial flow turbojet

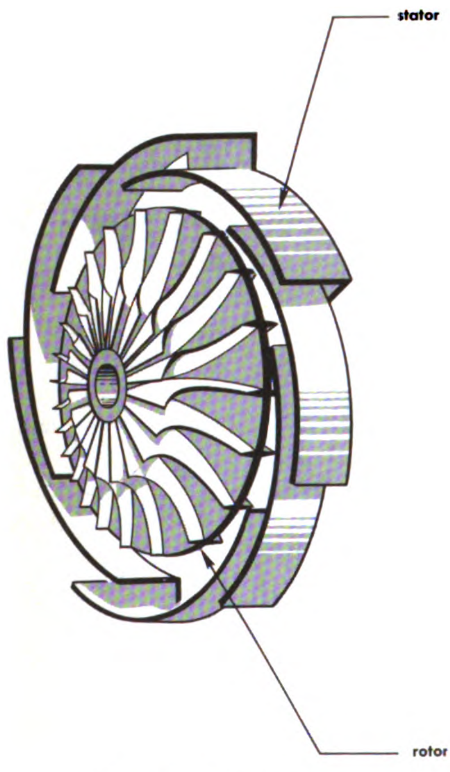


Attached accessories of turbojet engine

Compressor Section. The primary functions of the compressor section in turbojets are to receive, compress, and distribute the large masses of air to the combustion chambers.

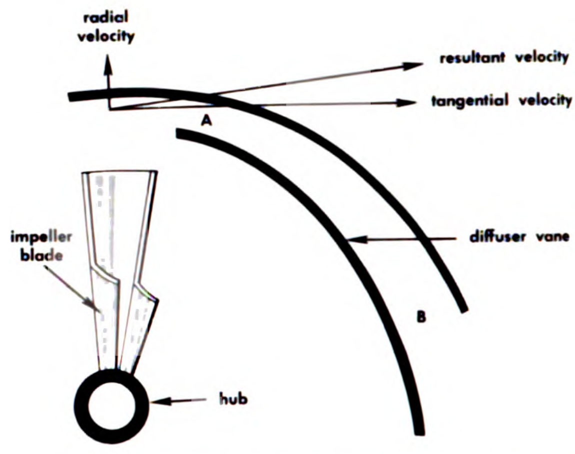
The *centrifugal compressor* illustrated in the diagram below consists mainly of a *stator*, often referred to as a *diffuser vane assembly*, and a *rotor* or *impeller*. The rotor consists of a series of blades which extend radially from the axis of rotation.

The double-faced type is commonly used, allowing air to enter on both sides. As the rotor revolves, air is drawn in, whirled around by the blades, and ejected at a high velocity created by centrifugal force.



Centrifugal-flow compressor

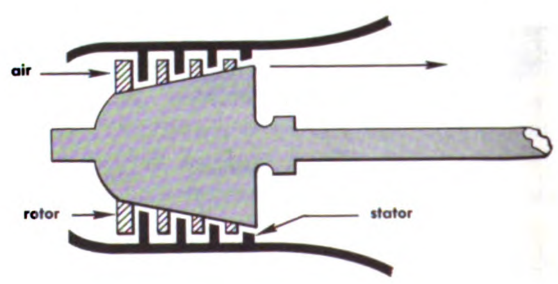
The stator consists of diffuser vanes which compress, as well as direct, the air into the various firing chambers. As the air leaves the impeller wheel, it has a large *resultant* velocity which directs the air into the diffuser vanes. The energy, which the masses of air acquire in the rotor as velocity, is converted into pressure, by the process of diffusion into a larger space. The air increases in pressure and decreases in velocity as it moves from A to B in the above diagram.



Sectional diagram of centrifugal-flow compressor

The *axial compressor* is similar to a propeller. The rotor again consists of a series of blades which are set at an angle and extend radially from the central axis. As the rotor of the axial compressor turns, the blades impart energy of motion, in both a *tangential* and *axial* direction, to the ram air entering through the front of the engine. A typical stator for an axial-flow compressor consists of a series of blades arranged in a circle around the inside wall of the compressor casing, immediately behind the rotor, and extending inward toward the central axis of the engine.

The stator *does not rotate.* The blades are set in an angular fashion, so as to turn the air thrown off the trailing edge of the first-stage rotor blades, and redirect it into the path of the second-stage rotor blades. *One rotor* and *one stator* constitute a *single-stage* compressor. A number of rotors and stators assembled alternately make up a *multistage* compressor.

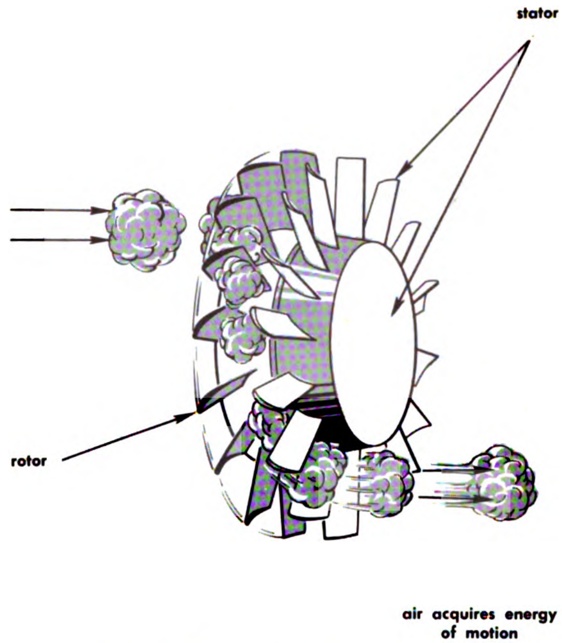


Four-stage axial-flow compressor

An example of a multistage compressor is illustrated at the bottom of the second column on the preceding page.

Air from the first row of compressor blades is accelerated and forced into a smaller space. The added velocity gives the air greater impact force. This impact forces air into a smaller space causing the air to become denser. This results in greater static pressure. The above cycle of events is continued through each stage. Therefore, by increasing the number of stages, the final pressure can be raised to any desired value. For visual purposes, the stator is shown in the diagram on the right with blades extending from a central drum outward. However, the most frequently used construction is with the blades extending *inward* as previously described.

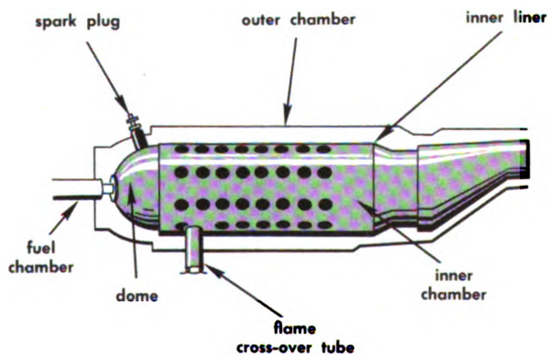
In both centrifugal- and axial-flow compressors, pressure ratio of the gases is given as exhaust pressure to intake pressure. As the speed of the compressor rotor is increased, the volume of the air passed by the compressor unit increases. At high-pressure ratios, the volume of air, passed for a given rotor speed, decreases. Above a given pressure ratio, the efficiency of operation for a given speed drops sharply. This drop is due to the development of pressure pulsations set up in the air as it passes through the compressor.



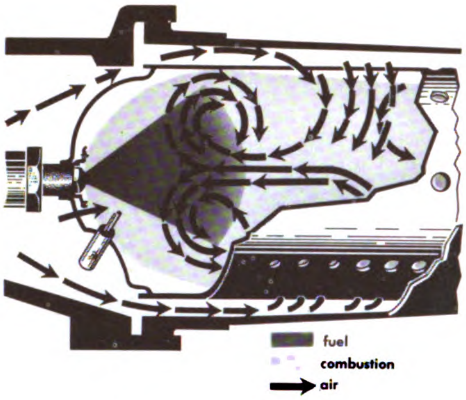
Single-stage axial-flow compressor

Combustion Section. The combustion section includes combustion chambers, spark plugs, a nozzle diaphragm, and a turbine wheel and shaft. The combustion chambers, or *burners*, in both types of turbojet engines have the same function and produce the same results. However, they do differ in size and number, depending upon the type of engine. One particular centrifugal-flow turbojet has 14 chambers with a spark plug located in chamber No. 7 and another in chamber No. 14. A widely used axial-flow engine has eight chambers with two spark plugs located in chambers directly opposite each other. In either case, each combustion chamber has the following parts: *outer combustion chamber, inner liner, inner-liner dome, flame cross-over tube, and fuel-injector nozzle*. The diagram on the right points out these components.

The outer combustion chamber serves to retain the air so that a high-pressure supply is available to the inner liner at all times. This air also serves as a *cooler jacket*. The inner liner houses the area in which fuel and air are mixed and burned. Many round holes in the inner liner allow the air to enter and mix with the fuel and high-temperature combustion gases. The forward end of the inner liner is allowed to slide over the dome to accommodate expansion and contraction. The aft ends of the burners are convergent to increase the velocity of the gases just before they pass through the nozzle diaphragm. The *flame cross-over tube* connects one chamber to the next, allowing ignition to occur in all chambers after the two chambers containing



Basic construction of turbojet combustion chamber



Propellant flow and combustion process in combustion chamber of turbojet

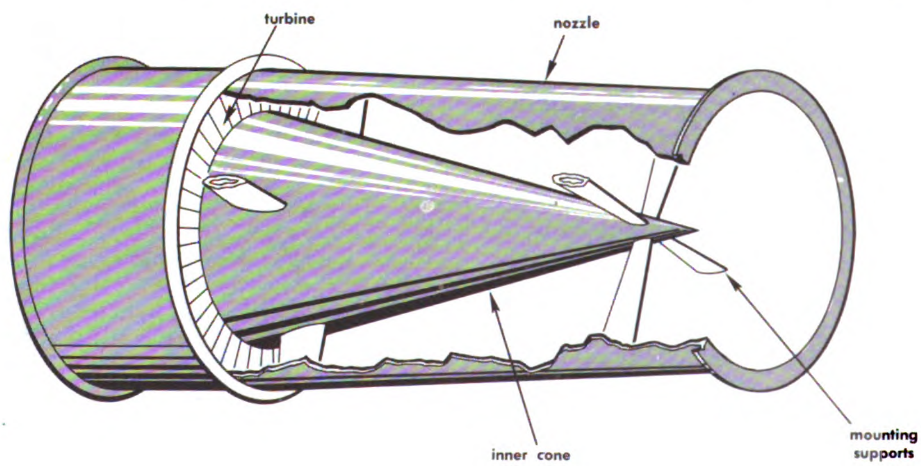
spark plugs have been fired. The above diagram illustrates the injection paths of fuel and air into the combustion chamber as well as the burning area.

Exhaust Section. The exhaust section, shown below, consists primarily of a *nozzle* and an *inner cone*. This assembly straightens out the turbulent flow of the exhaust gases caused by rotation of the *turbine wheel*, and conveys these gases to the nozzle outlet in a more perfect and concentrated gas-flow pattern.

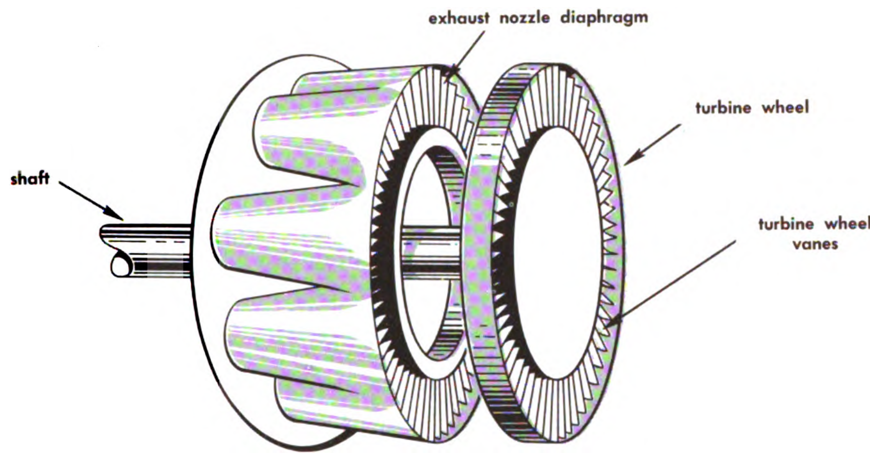
The *exhaust-nozzle diaphragm* is composed of a large number of curved blades standing perpendicular to the flow of combustion gases and arranged in a circle in front of the turbine wheel. By acting both as a restrictor and director, this diaphragm increases the gas velocity. Its primary function is to change the direction of the gases so that they strike the turbine-wheel vanes at, or near to, a 90-degree angle. This diaphragm *does not rotate*.

The impact of the high-velocity gases against the buckets of the turbine wheel causes the wheel to rotate. The turbine-wheel *shaft* is coupled to the compressor-rotor-assembly shaft. Thus, part of the energy of the exhaust-gases is transformed and transmitted through the shaft to operate the compressor and engine-driven accessories. Note the illustration at the right.

TURBOJET OPERATION. In general, turbojet operation may be summarized as follows: The rotor unit of the compressor is brought up to maximum allowable speed by the starter unit, which is geared to the compressor shaft for the starting phase. Air is drawn in from the outside, compressed, and directed to the combustion chambers. Fuel is injected through the fuel manifold under pressure and mixes with the air in the combustion chambers. Ignition occurs first in the chambers containing the spark plugs and then in the other chambers an instant later by way of the flame cross-over tubes. High-pressure combustion gases, plus the unburned *coolant* air, pass through the exhaust-nozzle diaphragm and strike the turbine blades at the most effective angle. The greater portion of the energy of the exhaust stream is absorbed by the turbine, resulting in a high rotational speed. The remainder is thrust. The turbine wheel transmits energy



Exhaust assembly of turbojet



Exhaust-nozzle diaphragm and turbine assembly

through the coupled turbine and compressor-rotor shafts to operate the compressor. Once started, combustion is continuous.

Afterburners. Afterburners were developed to satisfy special conditions in which large bursts of extra thrust are needed for a short period of time. For instance, an afterburner is needed for acceleration during takeoff runs and in climbs.

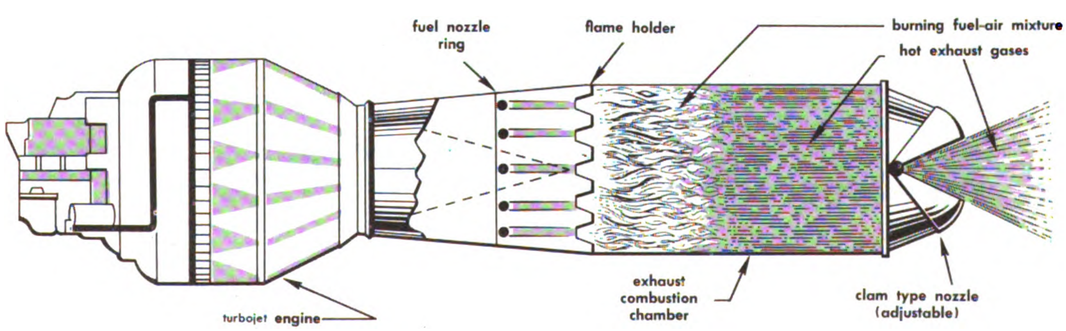
This additional thrust may be obtained by burning fuel in the tail-pipe section. That portion of air which served only as a coolant for the main combustion chambers is sufficient to support combustion of the additional fuel. The added thrust is large, but the overall efficiency of the turbojet *decreases* because the specific fuel consumption is greatly *increased*. The basic construction of a typical afterburner is shown below. Note the use of the clam-type exhaust nozzle, the area of which is adjustable to compensate for the changing flow characteristics that occur when the afterburner is cut-in. An afterburner boosts thrust nearly 30% for takeoff and from 70% to 120% when the missile is traveling in the neighborhood of 600 mph.

While air-jet engines are limited in area of operation because of their dependency upon the surrounding atmosphere for oxygen, rocket motors are not restricted by such conditions.

ROCKET MOTORS

Any missile for which the power for operation depends only on the mass and energy stored *within* the craft itself is called a *rocket*.

In rocket operation, chemical reaction occurs at a very rapid rate. This results in higher temperatures, higher operating pressures, and higher thrust development than in the case of the jet engines previously discussed. Con-



Turbojet after burner

trol over this action is desirable and is sometimes necessary to distribute the available energy over a longer period of time. Because of the high pressures which are developed in rocket motors, the convergent-divergent (DeLaval)-type nozzle is used so that more of the energy can be extracted from the gases after they have passed the throat section.

The basic principles involved in the action of other jet-propulsion units also apply to rockets.

Classification of Rocket Motors

Depending on the physical state of the propellant used, rockets are designated as either *solid* or *liquid* type.

A solid rocket is characterized by its short burning time, simple design, heavy construction, and nonintermittent operation. Therefore, it is primarily used as a booster unit or as a power plant for short-duration, high-speed missiles.

The liquid rocket unit may be thought of as a unit of longer-burning duration, relatively complicated design, with intermittent operation possibilities. This system has been widely used as a power plant for high-altitude, long range missiles. The V-2 and Viking were examples.

Solid Rockets

A solid rocket unit consists of the *propellant*, *combustion chamber*, *igniter*, and *exhaust nozzle*. A complete typical solid rocket motor is illustrated in the following drawing, which shows all of the components:

Propellants for solid rockets were discussed earlier in this chapter.

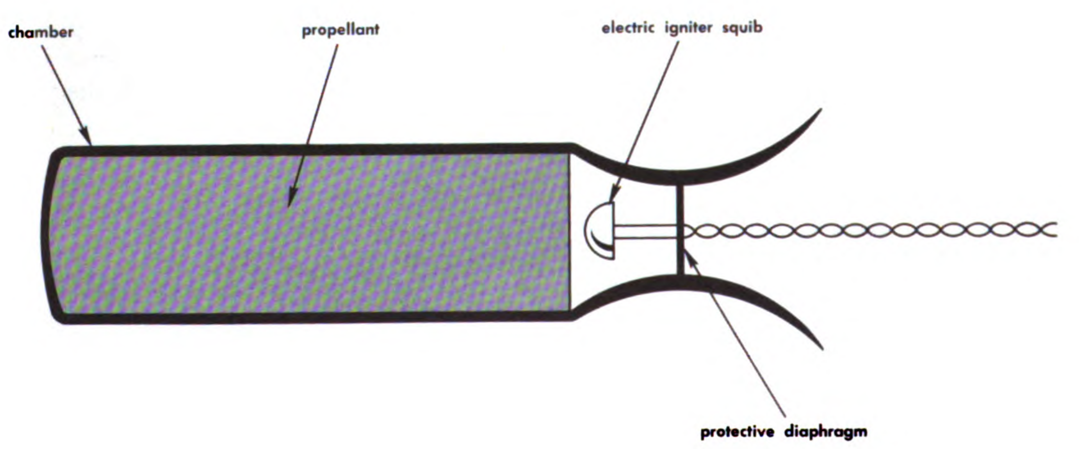
The combustion chamber of a solid rocket serves two purposes. First, it acts as a storage place for the propellant, and secondly, it serves as a chamber in which burning takes place. Depending upon the grain configuration used, this chamber may also contain a device for holding the grain in the desired position, a *trap* to prevent flying particles of the propellant from "clogging" the throat section, and *resonance rods* to absorb vibrations set up in the chamber.

The igniter consists of a small charge of black powder or some other substance that can be easily ignited by either *spark discharge* or a *hot wire* and that can produce a temperature high enough to ignite the main rocket propellant charge.

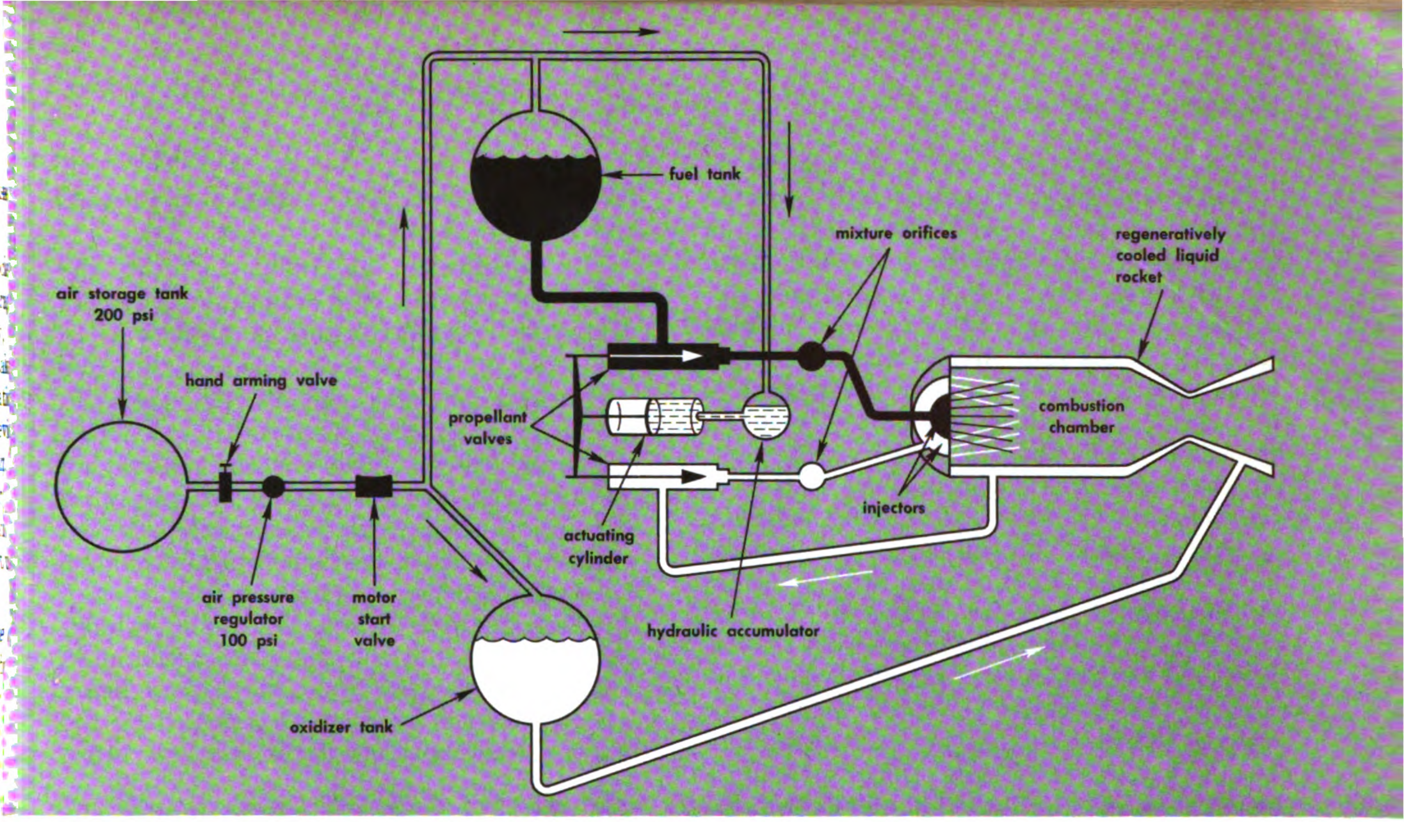
The exhaust nozzle serves the same purpose as in any other jet-propulsion system.

A solid-propellant rocket takes the same classification as the type of charge it uses, such as a *restricted-burning* rocket or *unrestricted-burning* rocket. The characteristics of each type were discussed in the solid-propellants section of this chapter. Because of previously mentioned characteristics of solid propellants, a solid rocket motor can be stored in the charged state.

The operation of a solid rocket is simple. After it is attached to a missile, the rocket is ignited electrically from a safe distrance. The igniter squib assembly is blown out and the



Components of solid rocket motor



Stored-pressure-feed system of liquid rocket

rocket burns continuously until the propellant supply is exhausted.

Cooling is not a great problem since the burning duration is short. One method, however, of preventing excessive heat from reaching chamber walls is in the use of a *hollow-restricted* charge. With this charge, burning takes place only on the inner surface, and the outer walls of the grain tend to act as an insulator. This becomes less and less effective as the grain is burned thinner and thinner. The exhaust nozzle is of heavy construction because of the critical temperatures in that section.

Liquid Rockets

The major components of a liquid-rocket system are the *propellant*, *propellant-feed system*, *combustion chamber*, *igniter* (if propellants are nonhyperbolic), and *exhaust nozzle*. A propellant-feed system is the only component with which you are not familiar.

PROPELLANT-FEED SYSTEMS. The commonly used types of propellant-feed systems are the *pressure-feed system* and the *pump-feed system*.

Pressure-Feed Systems. Pressure-feed systems are subdivided into *stored-pressure* systems and *generated-pressure* systems. In the case of a stored-pressure system, air or some other gas is stored under pressure within the missile before launching. It is injected, in controlled, amounts into the propellant storage tanks, causing a pressurized flow toward the combustion chamber. In a generated-pressure system, substances are carried within the missile to "generate" the high-pressure gas as needed. An example of such a substance is hydrogen peroxide which, when passed through a catalyst like sodium permanganate, decomposes to form a high-pressure vapor. This vapor is then injected into the propellant storage tanks.

Many other devices such as *valves*, *regulators*, *delivery tubes*, and *injectors* are necessary for the successful operation of either system.

The diagram at the top of this page shows the general relationship of the various major components of a *stored-pressure-feed system*. In the feed system shown, air is stored under a pressure of 2000 psi. The *hand-arming valve* is opened manually, just prior to launching. This operation allows the system to be pressurized up to the motor-start valve. The *air-pressure regulator* decreases the pressure to the desired value necessary for operation of the system components — in this case 100 psi.

The *motor-start valve* is electrically operated and opened from a safe distance after all personnel have cleared the immediate area. Air, at a pressure of 100 psi, enters and pressurizes the fuel and oxidizer tanks. These *propellant tanks* must be constructed of material that is not affected by the respective propellants. In addition, they must be strong enough to withstand the added air pressure. At the same time the propellant tanks are pressurized, air also enters the *hydraulic accumulator* and pressurizes the hydraulic fluid. The hydraulic fluid displaces the piston in the propellant-valve *actuating cylinder*, which in turn opens the propellant valves. Fuel and oxidizer, under pressure, now flow through the respective *mixture orifices*, which *regulate* the flow so that the correct mixture ratio is maintained. These orifices are simply restrictions in the line and are *flow-checked* prior to installation. In some cases the injectors perform this operation, in which case orifices are not necessary. The propellants are atomized by the injectors. Note that the oxidizer first circulates between the walls of the combustion chamber before passing through the cutoff valve. This is known as *regenerative cooling*.

Pressure-feed systems are used for powerplants in which the volume of propellants to be transferred is small and in which the weight rate of flow is low. In a system requiring enormous amounts of propellants that must be transferred at a high rate of flow, a much larger volume of gas would have to be stored at extremely high pressures. The tank necessary for such storage would then be large and space-consuming and of very heavy construction. These are undesirable features in a missile.

Pump-Feed Systems. Pump-feed systems are used with power plants designed to burn large volumes of propellants and with plants requiring a high weight rate of flow. A pump-feed system consists of a fuel pump and an oxidizer pump, driven by a turbine wheel. If the power for driving this turbine wheel is a gas generated by chemicals carried within the missile for that purpose, it is referred to as a *turbine-pump* system. If the turbine wheel receives its power from the exhaust gases of the rocket motor, it is called a *turbopump* system.

The diagram on the right illustrates the major components of a liquid-rocket power plant using a turbine-pump-feed system.

The stored air supply pressurizes the two tanks containing steam-generating agents. In the steam-generating chamber, the unstable compound "breaks down" in the presence of the catalytic agent, forming high-pressure steam. This steam drives the turbine wheel which in turn drives the two propellant pumps. The fuel and oxidizer are thus pumped through their respective feed lines and remotely controlled valves into the combustion chamber. The fuel is used as a coolant before being injected. Fuel in excess of that passed by the fuel injector flows through the overflow return line.

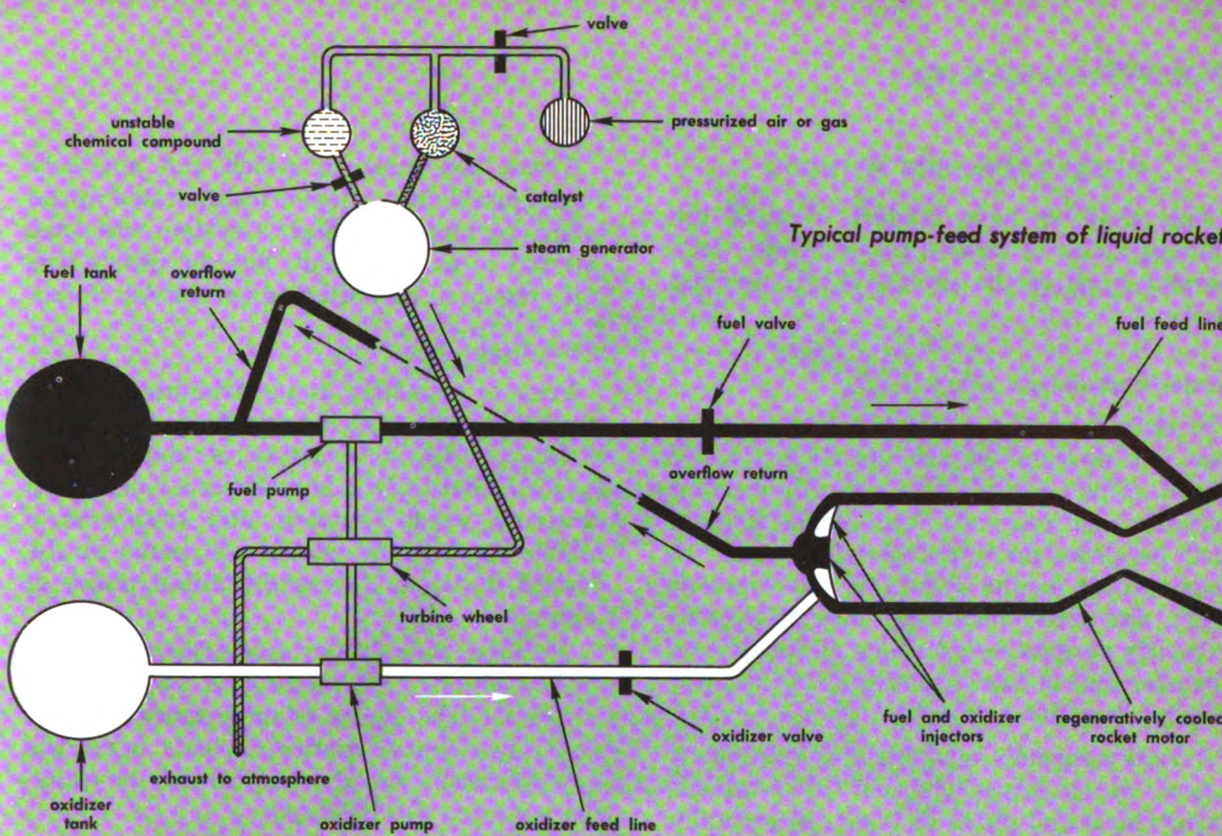
The turbopump-feed system would not use the steam-generating unit. Instead, the turbine wheel would be located partially in the rocket exhaust stream. Proper shaft and gear connections would carry the resulting rotational energy to propellant pumps.

MOTOR-COOLING SYSTEMS. Because of the intense heat developed in liquid-rocket combustion chambers, it is important that the inner walls of the chamber, throat, and exit be cooled. Uncooled operation over a prolonged period reduces physical strength and possibly would melt critical areas of the motor.

A *regenerative cooling* method is often used as shown in the two preceding liquid rocket diagrams. Before injection into the chamber, the fuel, or oxidizer, is circulated from front to rear between the walls. The heat absorbed by the fuel or oxidizer cools the chamber and adds to the energy originally contained in the propellant or oxidizer.

A *film cooling* procedure consists of low-velocity injection of a portion of the fuel, oxidizer, or some nonreactive liquid into the chamber at critical points. The fluid forms a protective film on the inner wall. The film absorbs heat and cools by evaporation.

A system using a *sweat cooling* method requires a combustion chamber whose inner wall is constructed of a porous material. The coolant enters through these pores, absorbing heat by the same evaporative process as in



film cooling. This method presents difficult problems. Porous construction tends to decrease physical strength. Also, lack of uniform pore distribution and variations in chamber pressure at various points make it impossible to control coolant flow.

Regenerative cooling, film cooling, and a combination of these two, are the most commonly used systems.

Since the field of guided missiles is relatively new, one can expect missile power plants to undergo frequent changes of design in the immediate future, resulting in greater efficiency of the units. However, this does not mean that present power plants will soon become obsolete. There is little question that the basic components and processes now found in the power systems will continue to exist in those systems for a long time.

Atomic Power for Missile Propulsion

Serious consideration is being given to the use of atomic power for propulsion of missiles Such a power plant would greatly increase the velocities and lengthen the operational periods of missiles. Present propulsion systems would become obsolete as major power plants. However, they may still serve as boosters for takeoff and acceleration during the first few miles of flight in order to prevent radioactive contamination of the launching area by an atomic powerplant.

One of the main advantages in the use of atomic power is that it provides an almost inexhaustible fuel or heat source. This advantage, in a field that is especially handicapped by high fuel-consumption problems, is most important. In any missile propelled by nuclear energy, the fuel supply would remain nearly constant. Enough fuel to start the reaction would be sufficient for sustained operation. However, some other "mass" must be carried within the missile to absorb the nuclear energy and be accelerated to generate thrust.

The major problems confronting the engineers are protecting the missile components and instruments from radiation damage and developing a nuclear power plant compact enough to be practical for missile propulsion.

Many years of extensive technical development may be necessary before nuclear energy is harnessed for use as a missile power plant. However, the outlook is promising, and the use of atomic energy will certainly make distance a minor factor in missile travel since large quantities of power would be attainable.

MISSILE LAUNCHING METHODS

Now that we have studied some of the different types of missile power plants, let's consider some of the problems of launching missiles.

Booster Assemblies

In most cases the power plant of a missile does not develop enough static thrust for takeoff. This condition is especially true of pulsejets and ramjets. Also, some missiles must be given an initial speed in order to insure proper engine operation, development of lift, and aerodynamic stability at takeoff. To satisfy various requirements of this nature, a booster unit is used in addition to the main missile propulsion system.

A *booster assembly* is an auxiliary propulsion system which imparts thrust to a missile during the initial phase of flight. It generally consists of a solid rocket motor and an attaching device. A solid rocket is best suited for this purpose because of its simple construction and operation and its ability to develop a high thrust in a short time. In some instances, a booster assembly includes an aerodynamic stabilizing surface. Usually a booster is attached directly to the missile.

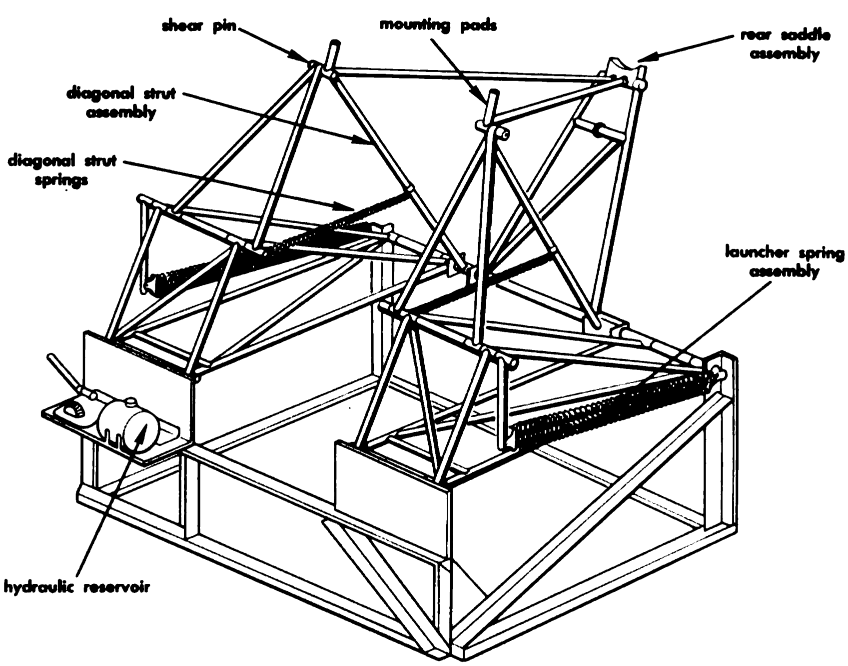
There are cases in which the auxiliary thrust-producing system is incorporated in the launching structure. Such a system was once referred to as a catapult, but the use of this term has been discontinued by the Air Force.

Special Launchers

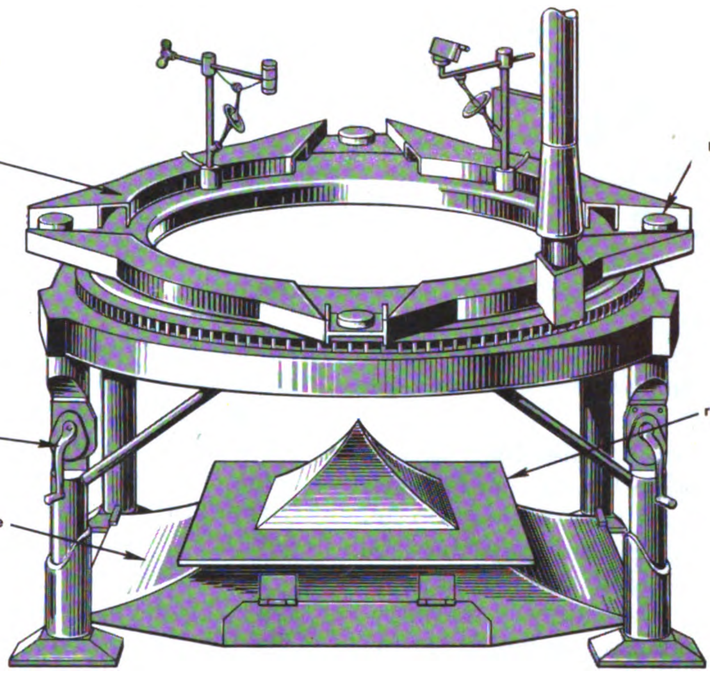
Prior to launching, some structure must be employed to support the missile in the desired attitude. Such a structure is known as a *launcher*. Basically, a launcher is a mechanical structure which *supports* and *constrains* a missile so that the missile will move in the desired direction during the initial phase of flight.

Mobility, *trainability*, and *high rate of fire* are desirable characteristics of missile launchers. The degree of importance of these and other characteristics depends upon the type of missile and the tactical use for which the launcher was designed.

ZERO-LENGTH LAUNCHER. A zero-length launcher exhibits no appreciable constraining action on a missile. A fraction of a second after initial motion begins, physical contact between the missile and launcher is broken. In order that the missile may attain flying speed in this short period of time, a powerful booster assembly must be attached to the missile.



Zero-length launcher



Platform launcher

Notice the diagram to the left of a zero-length launcher. The wings of a missile are supported by the two mounting pads, and the tail rests on the rear saddle assembly. The hydraulic system provides a means for raising the launcher to the desired angle. The side view of this launcher presents a parallelogram with a diagonal strut attached to the mounting-pad junction by a shear pin. The same holds true for the opposite side. When the thrust developed by the booster and missile power plant is adequate to insure flying speed, the diagonal strut pins are sheared, allowing the parallelogram to collapse in the forward direction and permitting the missile tail and booster assembly to pass. The booster assembly automatically detaches itself from the missile when its propellant charge is exhausted.

PLATFORM LAUNCHER. A platform launcher supports a missile in a vertical position. Such launchers are used with large, rocket-powered, long-range missiles of the V-2 type. No booster unit is necessary since the powerful rocket motors develop adequate takeoff thrust. Proper attitude during the low-velocity period must be maintained by some automatic system because of the ineffectiveness of aerodynamic control surfaces.

The platform launcher is simple in construction and trainable through a few degrees of tilt. The drawing at the top of this page illustrates the construction of such a launcher. Built-in jacks provide a means of raising, lowering, and leveling the table. A platform is provided for use by the operators during *set-up* and *check-out* procedures. This platform must be removed before launching. A blast plate deflects the concentrated exhaust stream and prevents violent erosion of the area beneath the launcher during takeoff.

RAMP LAUNCHER. A ramp launcher consists of two rails attached to the top surface of a ramp which can be of either fixed or adjustable angle of elevation. Shown on the next page is a ramp launcher with a missile in position. Because of the massive size of this type of launcher, its construction with trainable capabilities presents a difficult problem. The V-1 ramp launcher was 150 feet in length and tapered from zero to 16 feet at the high end. A number of variations in the source of auxiliary thrust units have been utilized throughout the history of ramp launchers. In some cases a sled and solid-rocket booster unit were attached to the bottom surface of the missile. It accelerated along the rails attaining a velocity of approxi-

Ramp launcher

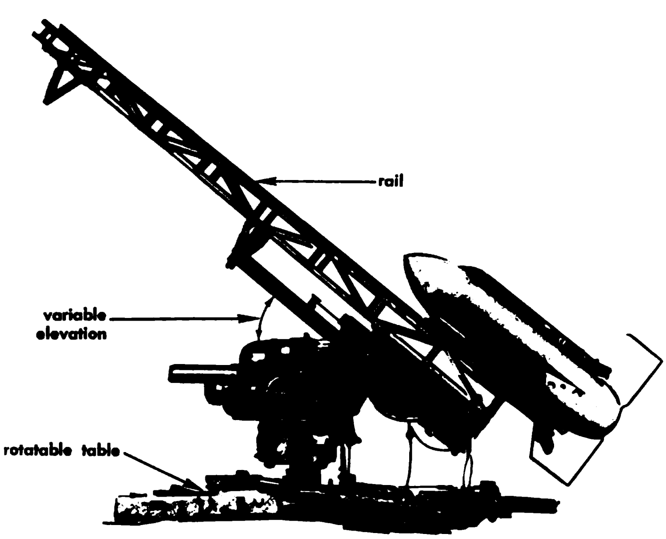
mately 200 mph near the end of the ramp. The booster unit dropped off after about 400 feet of climb.

High-pressure steam has also been used as a source of auxiliary thrust. A tubular chamber was placed between the rails running the full length of the ramp. A piston inside the tube was connected to the missile. High-pressure steam, generated by the catalytic decomposition of hydrogen peroxide, drove the piston which in turn accelerated the missile along the rails.

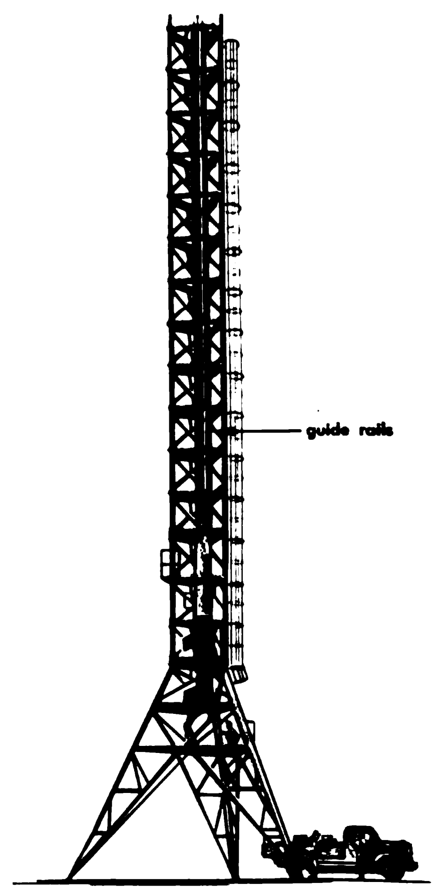
Another method of developing auxiliary thrust was by burning powder acrtridges mounted on the side of the launcher. Here again the thrust-producing unit stays with the launching structure when the missile takes off.

MONORAIL LAUNCHER. The monorail launcher illustrated just below can be elevated to an angle of approximately 90 degrees and rotated through 360 degrees. It is an ideal surface-to-air launcher because of of its *trainability* and *rapid-rate-of-fire* capabilities. Its short rail length limits use to missiles with powerful rocket-propulsion systems. Contact between missile and rail is maintained by use of a slotted attaching device.

A variation of this launcher consists of replacing the single rail with a slotted trough. The bottom surface of the missile is then equipped with runners to decrease friction and to constrain the missile.



Monorail launcher

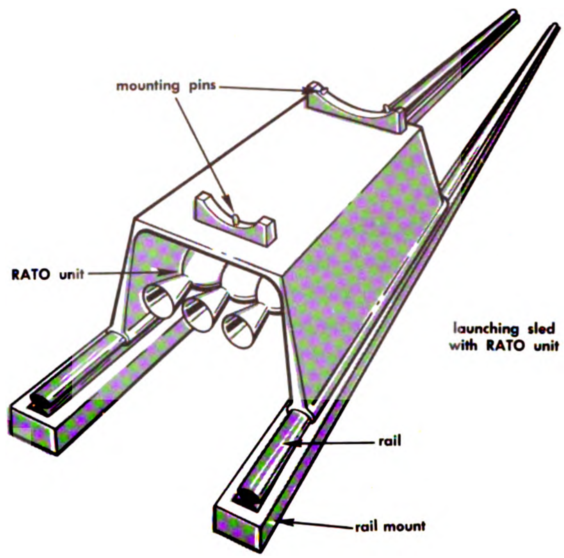


Tower launcher

TOWER LAUNCHER. Note the drawing of a tower launcher on the preceding page. This type of launcher has a derrick-type construction which holds the missile in a vertical, or near vertical, position during the launching phase. Some tower launchers are trainable through a few degrees of tilt in order to compensate for wind or other factors which may affect the desired trajectory. They must have enough height to allow the missile to attain adequate speed for aerodynamic stability before breaking contact with the tower. The major use of tower launchers is for high-altitude rockets which fly a ballistic-type trajectory.

SLED LAUNCHER. A sled launcher utilizes a launching sled equipped with solid rockets for development of thrust. The sled and the missile which it supports are propelled along two rails which resemble railroad tracks. After the missile is airborne, the launching sled is stopped by a water braking system and returned to the starting point for subsequent launchings.

The illustration at the upper right is intended to give a general conception of sled, rail, and assisting propulsion-unit relationships. Many details have been omitted to avoid classification problems. Actually the rails must be carefully leveled, and they may be built to extend a mile or more in length. The braking process, not shown in the illustration, is an elaborate arrangement.



Sled launcher

Multistage Launching

Multistage launching, as shown in the illustration on the following page, consists of two or more missiles, or missile sections, each containing a separate propulsion system. These sections are called "stages." As an example, in a three-stage system, the first stage develops takeoff thrust and propels the entire missile system until maximum velocity is attained. At this point the first stage falls away and the second-stage power plant ignites automatically, increasing the velocity already attained to a new maximum. The second stage is then disconnected, and the third stage further increases the velocity imparted to it by the two previous stages. In this way, much higher velocities and greater altitudes are reached than would be possible with a single-stage missile using one propulsion system. The Wac Corporal was successfully launched from the nose of a V-2 (two-stage "Project Bumper"), enabling the Corporal to reach an altitude of 250 miles.

With present-day propellants, multistage launching seems to be the most practical means of attaining the theoretical escape velocity from the earth and making space travel a reality.

MISSILE PROPULSION SYSTEMS IN BRIEF

Before looking into the physics involved in the design of guided missiles, let's briefly recap some points brought out in this chapter.

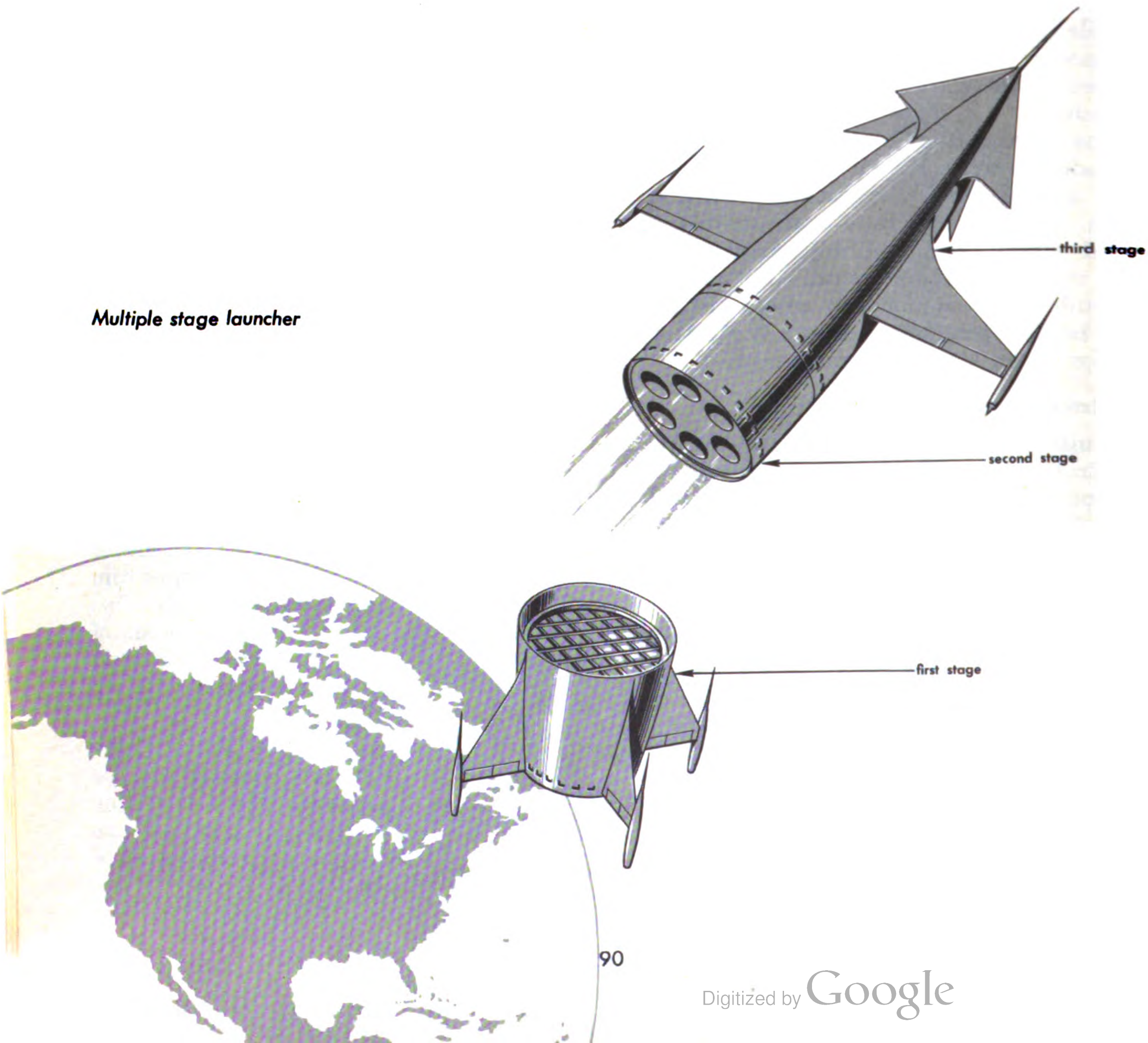
You read that jet propulsion is a means of locomotion brought about by the momentum of matter from within the propelled body. This thrust is obtained through the combustion of a solid propellant, such as ballistite, or a liquid propellant, such an aniline mixed with an oxidizer.

The basic components of a jet-propulsion system are the combustion chamber and the exhaust nozzle. In many missiles, depending on the propellant used, injectors and an ignition system are important components of each combustion chamber.

Jet-propulsion systems can be classified into air-jet engines and rocket motors, the air-jet engines being dependent upon the surrounding atmosphere for oxygen while rockets are not dependent on the atmosphere. Ramjets pulsejets, shrouded pulsejets, and turbojet engines comprise the air-jet engines of the present day.

The possibilities of using atomic power for missile propulsion are imminent. As this chapter stated, the present propulsion systems would become obsolete as the major power plants if atomic energy were to be utilized. Even so, present systems more than likely would be of continued use as boosters. This use of present systems as boosters would prevent radioactive contamination of the launching areas by the atomic power plants. The main advantage of atomic power plants would be the almost inexhaustible energy source.

Launching of a missile presents special problems. In most cases the power plant of a missile does not develop enough static thrust for takeoff. Also, some missiles, such as ramjets, require an initial speed before their engines can operate properly. To meet these problems, booster units, usually solid rockets, are used in addition to the main propulsion systems. Launchers are used to support and constrain missiles so that the missile will move in the desired direction during the initial phase of flight.



Multiple stage launcher

Chapter 4

Physics Involved in Guided Missile Design

You will find that the configuration of missiles and the design of all components used in their fabrication, as well as the systems used for their guidance and control, are all governed by basic laws and principles of physics.

The function of each electronic or mechanical component of a missile involves some physical principle and/or property which has been found applicable to the specific requirement. Many of the components are simple in form and perform a single fundamental operation; other components may operate simultaneously or in conjunction with other simple components to perform highly complex operations, as in the case of the components and systems used for missile guidance, control, and instrumentation.

In this chapter you will become familiar with physics as it applies to missile design, and you will learn how the function of electronic or mechanical components involves physical principles and/or properties. The chapter also explains how optical and electronic principles work together. It discusses the physics of transistors and the final section of the chapter goes into considerable detail on the modulation of carrier waves. A knowledge of the material covered in this chapter will help you to understand the functions of components and systems employed in missile operation. The next chapter begins the discussions of such components and systems.

A REVIEW OF NEWTON'S LAWS OF MOTION

Newton's laws of motion are specific examples of basic laws of physics which are involved in guided missile operations. Even though you are already familiar with these laws, they are important enough to the study of missiles to be repeated here.

Newton's first law of motion states that "A body at rest remains at rest, and a body in motion continues to move at constant speed along a straight line unless the body is acted upon in either case by an unbalanced force."

Newton's second low of motion states that, "An unbalanced force acting on a body causes the body to accelerate in the direction of the force, and the acceleration is directly proportional to the unbalanced force and inversely proportional to the mass of the body." This law in formula reads:

$$a = \frac{F,}{m} \text{ where } a = \text{acceleration of the body}$$

F = the unbalanced force acting upon the body, and

m = the mass of the body

Newton's third law of motion states that "For every action, there is an equal and opposite reaction, and the two are along the same straight line." The diagram below helps clarify the meaning of this third law:

$$A \underset{\text{Reaction}}{\overset{\text{Action}}{\rightleftarrows}} B$$

In stating this law, the term "action" refers to the force which one body exerts on a second body, and "reaction" refers to the force which the second body exerts on the first.

The principles stated in the foregoing paragraphs are involved in missile design and operation. To control the flight of the missile, we must employ devices to act with or to react against the natural forces present along the flight path of the missile.

In some cases, it is possible to cause one undesired force to react against another in order to cancel or nullify the effects of undesired forces. Such operations may become complex, depending upon the number of interacting forces involved. In cases involving a considerable number of interacting forces, complex systems, such as those employed in missile navigation, are needed. Some forces that missile navigation systems must deal with are gravity, inertia, acceleration, Coriolis effect, and magnetism.

PHYSICAL PRINCIPLES AND PROPERTIES INVOLVED IN MISSILE OPERATION

To understand fully the functions of the basic components and systems employed in missile operation, you should have further information pertaining to the physical principles and properties involved. You need to be familiar with such terms as inertia, gravity, circular motion, centripetal and centrifugal force, rotational motion, radius of gyration, and motion of precession.

These terms and others are discussed in the following paragraphs. It is neither practical nor desirable to present complete analyses of the meanings of the terms here. Instead, you will be given the relationships among the various physical principles and properties to which the terms refer. You need to know these relationships so that you may more readily recognize those which may be involved in the basic units and operational systems of missiles.

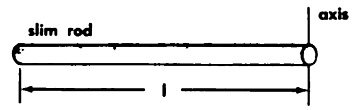
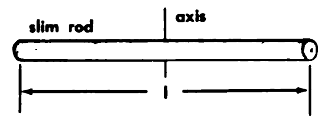
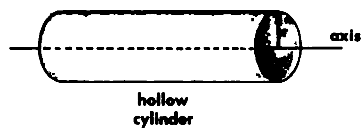
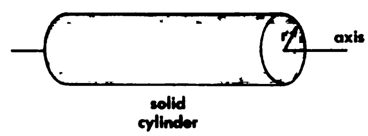
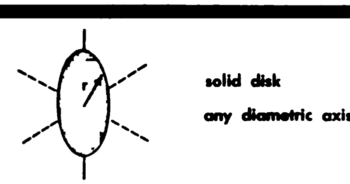
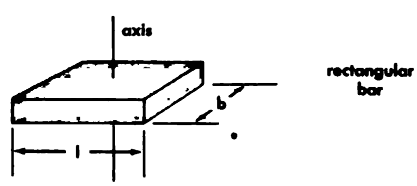
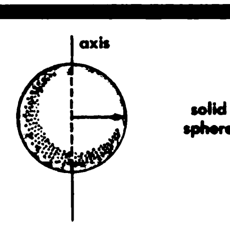
Inertia

Regardless of the specific purpose or design of a missile component and regardless of whether it be electronic or purely mechanical, the operation of the component is dependent upon one or more physical properties. One physical property inherently common to matter is *inertia.*

Inertia is defined as a property of matter by which it will tend to remain at rest, or if in motion will tend to follow the same straight line or direction unless acted upon by some external force. This definition is stated in Newton's first law of motion. You may also think of inertia as the opposition offered by a body to any change of motion; that is, an unbalanced force is required to give it linear acceleration.

The property of inertia implies a broader concept to the definition of *mass* than mere quantity of matter. You may think of mass as the property of an object by virtue of which it possesses inertia.

MOMENT OF INTERTIA. A tendency, or measure of a tendency, to produce motion about an axis is referred to as a *moment.* A *moment of inertia* (I) of a body is equal to the mass of one part of the body multiplied by the square of its distance from the axis of rotation, plus the mass of another particle multiplied by the square of its distance from the axis of rotation, plus — and so on, until all the particles of the body have been included. In other words, the masses of the particles of which a body is composed and squares of the individual radii that extend from these particles to the axis determine the moment of inertia of the body.

SHAPE	(I)	
slim rod of mass (M) and length (l), about a transverse axis through one end.	$\frac{Ml^2}{3}$	
same, but with transverse axis through center.	$\frac{Ml^2}{12}$	
hollow cylinder (thin wall) of mass (M), radius (r) and of any axial length, about its own axis.	Mr^2	
solid cylinder or disk of mass (M), radius (r), and of any axial length, about its own axis.	$\frac{Mr^2}{2}$	
solid disk of mass (M), radius (r), about any diameter.	$\frac{Mr^2}{4}$	
rectangular bar of mass (M), length (l), and width (b), about an axis through its center and at right angles to dimensions b and l.	$M\frac{b^2+l^2}{12}$	
solid sphere of mass (M), and radius (r), about any diameter.	$\frac{2Mr^2}{5}$	

Moments of inertia of various geometric shapes

Some useful values of the moments of inertia of various geometric shapes are given in the table above.

Moment of inertia is often expressed in terms of the aggregate mass of the body and a single fictitious radius called the *radius of*

gyration. The radius of gyration of a body is defined as the distance from the body's axis of rotation to a point at which the entire mass of the body may be considered concentrated without altering the moment of inertia.

Velocity

You already know that velocity is a representation of speed and a specific direction of motion. The velocity of an object may be constant or varying. If velocity is varying, the variation may be in speed or in direction or both. When variation takes place in direction, the term angular velocity is applied to describe the motion. Angular velocity is the rate at which the object changes in direction.

The angular velocity of a rotating body may be represented by an axial line having a length which indicates the numerical value of the velocity.

The direction of the vector, as shown in the figure above right, represents the direction in which an ordinary right-hand threaded screw would travel if turned in the direction in which the body (wheel) rotates.

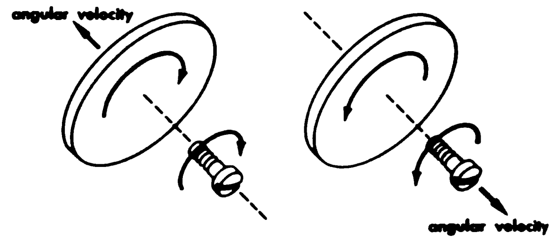
When a rotating body is given an additional angular velocity about the same axis, the resulting velocity will be their algebraic sum. For example, if a body rotating at 10 radians per second in a clockwise direction is given an additional angular velocity of 2 radians per second in a counterclockwise direction, its resulting velocity would be 8 radians per second clockwise about the same axis. However, if the added velocity is about a different axis than that of the original rotation, the resulting velocity must be obtained by vector addition.

Radians are a means of measuring angles based upon the relative dimensions of a sector of a circle. Radians show a ratio of an arc length to the radius.

Acceleration

Acceleration is the time rate of change of velocity. It represents motion in which the velocity changes from point to point.

When the velocity of a mass moving in a straight line changes by equal amounts in equal intervals of time, the acceleration is said to be constant and the motion *accelerated uniformly*.



Angular velocity as a vector

Moving objects rarely exhibit constant acceleration; however, a free-falling body in a vacuum near the surface of the earth is one example of uniformly accelerated motion. Experiment has shown that this constant acceleration resulting from gravity is 32 ft/sec². It is represented in physics formulas by the letter g.

When acceleration is not known to be constant, the average acceleration may be expressed by the equation:

$$a = \frac{V_f - V_o}{t - t_o}, \text{ where}$$

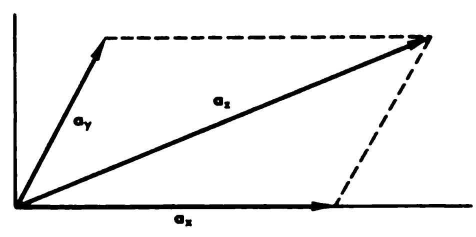
V_f = final velocity

V_o = initial velocity

$t - t_o$ = time interval during which velocity changed from V_o to V_f

The units of acceleration depend upon the unit of velocity as well as upon the unit of time during which the velocity changes.

Based on the above equation, acceleration is equal to the difference of two vectors (velocities) divided by the scalar quantity (time) and is a vector quantity which may be determined as the resultant by the common parallelogram method of vector addition. For example, if a mass is subjected to accelerations a_x and a_y, the resultant acceleration a_z may be obtained as shown in the figure below. The dotted lines are extended along a path



Vector resultant of two accellerations

opposite and parallel to a_x and a_y. The point at which the dotted lines intersect determines the resulting acceleration, or resultant.

Note that acceleration is not the time rate of change of *speed;* it is the time rate of change of velocity treated as a vector quantity. A mass moving with constant speed in a circle is accelerated despite the fact that the time rate of change of speed is zero. The velocity of the mass is changing continuously since the direction of motion is changing. Thus, the time rate of change of velocity is not zero. Such acceleration is termed normal acceleration and is discussed later in this chapter with reference to circular motion.

For any kind of motion, the distance (s) of an object from its starting point is obtained as the product of average velocity (V_{av}) and time (t).

$$S = V_{av}t$$

When the motion is uniformly accelerated (a is constant), the relationship between the average velocity (V_{av}) and the initial velocity (V_o) and final velocity (V_f) may be expressed by the equation

$$V_{av} = \frac{V_o + V_f}{2}$$

because in terms of initial and final values the average of any two quantities is equal to one-half their sum. By substituting and transposing from the equation

$$a = \frac{V_f - V_o}{t - t_o}$$

and the equation

$$V_{av} = \frac{V_o + V_f}{2}$$

to the equation

$$S = V_{av}t \text{ or } \left(V_{av} = \frac{S}{t}\right)$$

you can derive the equation

$$S = V_o\, t + \frac{at^2}{2}$$

Here's how the latter equation is derived:

From

$$a = \frac{V_t - V_o}{t - t_o}$$

in which $t_o = 0$,
you get $V_f = V_o + at$.

If in equation

$$S = V_{av}\, t$$

you substitute for V_{av} its value

$$\frac{V_o + V_f,}{2}$$

you obtain

$$S = \frac{(V_o + V_f)\, t}{2}$$

Combining this equation with equation

$$V_f = V_o + at$$

will result in equation

$$S = \frac{(V_o + V_o + at)\, t}{2}$$

or

$$S = \frac{(2V_o + at)\, t}{2}$$

By further reduction you obtain

$$S = V_o t + \frac{at^2}{2}$$

which is the original formula.

From this equation the average acceleration (a) required to displace an object a distance "s" in "t" seconds, assuming zero initial velocity, is found to be:

$$S = \frac{1}{2}\, at^2$$

where S = distance in feet
a = acceleration in ft/sec/sec
t = time in seconds

If initial velocity (V_o) is zero, the term

$$V_o\, t = o$$

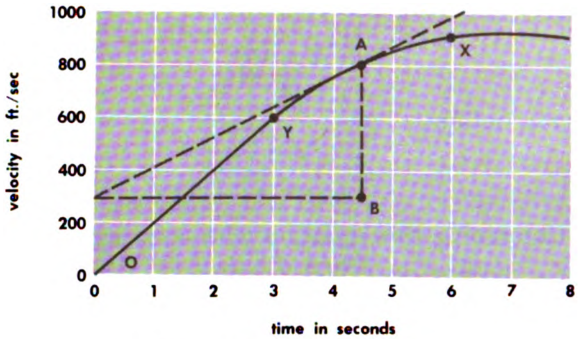
and equation

$$S = V_o\, t + \frac{at^2}{2}$$

reduces to

$$S = \frac{at^2}{2} \text{ or } S = \frac{1}{2} at^2$$

A practical example of the application of these formulas pertaining to accelerated motion may be illustrated by assuming that a rocket is launched from a position of rest, and it attains a velocity of 200 feet per second during the next second, and so on. Thus, it can be said that the rocket has an acceleration of 200 feet per second in each second. By the end of three seconds, the rocket will have a velocity of 600 feet per second, and so on, as long as it can maintain the same acceleration.



Acceleration curve of a rocket

During this period of increasing velocity, every value of velocity from zero up to the maximum velocity attained by the rocket is passed through at some instant of time.

If you plotted a graph of the instantaneous values of velocity, you would find that they would describe a curve as shown in the graph above.

The curve is straight from "O" to "Y," showing that during the first three seconds the rocket gains velocity at a uniform rate; that is, the acceleration is constant.

Beyond "Y," the curve slopes off indicating a reduction in acceleration until it becomes horizontal at "X," indicating zero acceleration or constant velocity.

If the acceleration of a body is considered as the change of its velocity during any interval of time divided by the duration of that interval, the definition is based upon change of velocity and not upon distance traveled. The definition may be stated in the form of an equation by supposing the motion of the body to be observed for some specific interval of time.

Taking "V_o" as the initial velocity of the body at the beginning of the period of observation and "V_f" as its final velocity at the end of the period, the change in velocity is

$$V_f - V_o$$

And if this change occurs in an interval of time t, the average acceleration of the body throughout that interval is

$$av = \frac{V_f - V_o}{t}$$

Referring to the figure showing the acceleration curve of a rocket, the velocity of the rocket under observation increases from 600 to 900 feet per second during the time interval from three seconds to six seconds after launching. But during this period the velocity does not increase uniformly as it does in the period from the instant of launching to three seconds after launching; therefore, the acceleration varies from instant to instant beyond point "Y".

However, by using the formula

$$a_{av} = \frac{V_f - V_o}{t}$$

the average acceleration of the rocket may be computed over the interval of time between the third to the sixth second, as follows:

$$a_{av} = \frac{900 \text{ ft/sec} - 600 \text{ ft/sec}}{6 \text{ sec} \quad - \quad 3 \text{ sec}} = \frac{300 \text{ ft/sec}}{3 \text{ sec}}$$

$$= 300 \text{ ft/sec} \times \frac{1}{3} \text{ sec}$$

$$= 100 \text{ ft/sec}^2$$

Similarly, by choosing any time interval during the first three-second period after launch (period of uniform acceleration), a result of 200 ft/sec² for the acceleration is obtained.

Referring again to the curve, if you solve for the average acceleration throughout the time interval between ½ second after launching (V_o) and two seconds after launching (V_f), you will find:

$V_o = 100$ ft/sec, $V_f = 400$ ft/sec, $t = 1\frac{1}{2}$ sec $(2 - \frac{1}{2})$.

Using these values in the formula, the equation becomes:

(400 ft/sec − 100 ft/sec) ÷ (1½ sec)
or (300 ft/sec) ÷ (1½ sec) = 200 ft/sec²

As the time interval is made shorter and shorter, the average acceleration approaches nearer and nearer to the instantaneous acceleration.

In the limit, for an infinitesimal change of velocity "dv" (delta v) occurring in an infinitesimal time interval "dt" (delta t), the instantaneous acceleration "a" may be expressed as $a = \frac{dv}{dt}$.

Concisely defined, "Acceleration is the time rate of change of velocity," and its value at any specific instant is represented graphically by the slope of the velocity-time curve at the

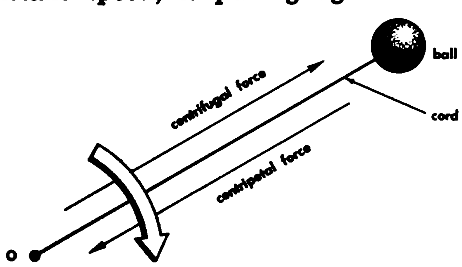
corresponding point. Thus, from the figure at left of the acceleration curve of a rocket, you can find the instantaneous acceleration at any time after launching by drawing a tangent to the curve at the point where it crosses the specific time ordinate and determining the slope of the tangent. For example, if you wish to find the instantaneous velocity of the rocket computed on the acceleration curve at 4½ seconds after launching, you can draw a tangent to the curve at point A, where the curve crosses the 4½-second time ordinate. Since AB = 800 ft/sec − 300 ft/sec or 500 ft/sec, the slope is 500 ft/sec ÷ 4½ sec = 111 1/9 ft/sec².

This value is the instantaneous acceleration of the rocket at the specific time of 4½ seconds after launching.

Centrifugal and centripetal forces. In accordance with Newton's second law, a body in motion will not deviate from a straight line unless a lateral force is exerted upon it. When a locomotive arrives at a curve in the track, its forward motion causes the flanges on its wheels to press outwardly against the edge of the outer rail; and, conversely, the rail presses inwardly against the flanges. The locomotive changes direction as result of this inwardly directed force and follows the curved track.

The lateral force required for motion along a curve also may be illustrated by a ball attached to a cord, as shown below. When the ball is whirled around at the end of the cord, the ball pulls outwardly, exerting a *centrifugal force* on the cord. The result of the centrifugal force is to cause the cord to become taut and exert an inward pull, *centripetal force,* upon the ball.

The ball, traveling along a circular path at constant speed, is pulling against the cord



Forces acting on a ball whirled around at end of cord

with a constant force directed outwardly along a radial line from the center of its path of rotation. Since for every action there is always an equal and opposite reaction, the cord exerts an equal inward force upon the moving body.

Thus, any motion along a curve involves *centrifugal* and *centripetal* forces which are equal and opposite, and both forces are exerted in the plane in which the curve lies. However, although equal, these forces cannot balance each other because they are not exerted upon the same object.

An unbalanced force always produces acceleration. The centripetal force acting upon a body in circular motion continually accelerates it toward the center of the circle. The body moves in a circle as a result of this inward motion combined with its forward motion.

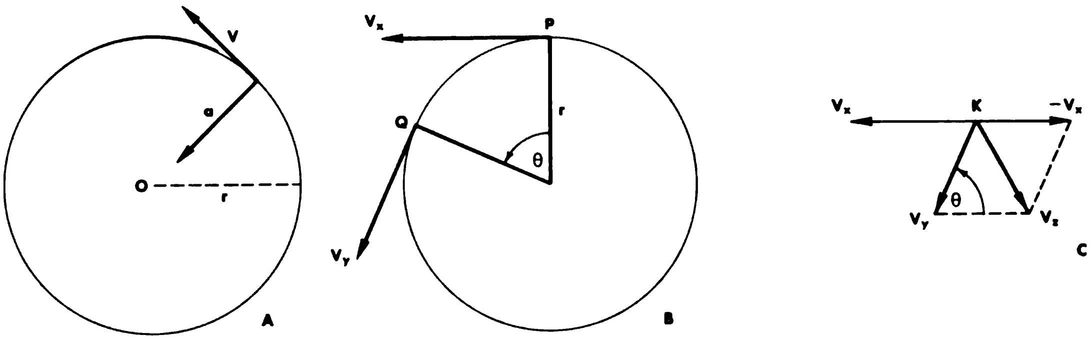
Note again the fact that when a body follows a circular path at constant speed, *its velocity is changing continually*, not in magnitude but in direction.

Assume a body to be moving at a constant speed "v" around a circle of radius "r" which is centered at "O," as shown in diagram "A" on next page. Assume also that in an interval of time "*t*" it moves a distance PQ = vt, as shown in diagram "B". Its velocities at "P" and "Q" may be shown by vectors "V_x" and "V_y" tangent to the circle at points "P" and "Q" respectively. The velocities in magnitude are equal to "v."

These velocities differ in direction, indicating that some additional velocity must have been given to the body in moving from "P" to "Q" to change its velocity from "V_x" to "V_y." To determine this additional velocity, "V_x" and "V_y" can be drawn from a common point "K," as in diagram "C" on next page. "V_x" is subtracted from "V_y." This subtraction is done by reversing the direction of "V_x" and adding in accordance with the parallelogram method of vector addition as explained on page 94. As indicated, the change in velocity is "V_z."

As the velocity change from "V_x" to "V_y" occurs during time t, the average acceleration throughout time "t" is:

$$a = \frac{V_z}{t}$$



Determining acceleration of a body moving in a circle

To determine the instantaneous acceleration, first notice that the angles designated θ (theta) are equal. Now, by assuming progressively shorter time intervals, the angles indicated by θ would simultaneously become progressively smaller. Thus, when zero time as a limit is approached, the sector POQ in figure "B" more nearly resembles the isosceles triangle, the equal sides of which form the angle θ of figure "C." This resemblance increases until finally:

$$V_z/V_y = PQ/OP$$

or

$$at/V_y = vt/r$$

Since $V_y = v$ numerically, you can derive the magnitude of the acceleration "a" by substituting "v" in the second term of the equation "$at/V_y = vt/r$." You get:

$$at/v = vt/r$$

By cross-multiplying, you get:

$$atr = v^2t$$

After cancelling the t's, you have:

$$ar = v^2$$

Now by transposing r, you get the simplified equation:

$$a = v^2/r$$

The above process shows that acceleration is a vector quantity. The direction of this vector quantity can be found by noting that the shorter the time interval taken, the more nearly "V_z" becomes perpendicular to "V_x" and "V_y," until at the limit (t = o), it is perpendicular to both vectors "V_x" and "V_y." In this manner you show that the *centripetal acceleration is directed toward the center of the circle.*

The equation

$$F = \frac{W\ a}{g}$$

in which "a" is the acceleration caused by an unbalanced force "F" acting on a body of weight (W) and "g" is the acceleration due to gravity, may be used to show the force that must be exerted upon the body to produce the acceleration.

Therefore, the centripetal force acting upon a body of "W" pounds when moving at speed "v" feet per second around a curve of radius "r" feet may be expressed:

$$F = \frac{W\ v^2}{g\ r}$$

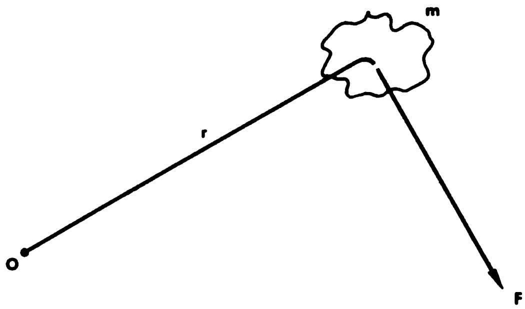
Note from this expression that the centripetal force acting on a body which is moving in a circular path varies directly as the square of the speed and inversely as the radius of the circle.

The force with which a body is attracted toward the earth is diminished as a result of the rotation of the earth. The effect of this rotation is most noticeable at the equator where velocity resulting from the earth's rotation is greatest. This influence of gravity is about 0.0035 or seven pounds per ton at the equator. Velocity is zero at the poles. An object which weighs 2000 lbs at the equator weighs about 2007 lbs at the poles.

Angular Acceleration

For purposes of illustrating the combined action of forces, consider how angular acceleration is produced by torque.

Torque (a moment of forces) is a rotational effect on a body and is measured by the



Angular acceleration produced by torque

product of the force present and the perpendicular distance from the axis of rotation to the line of action of the force. The perpendicular distance is referred to as the *lever arm;* therefore, torque may be expressed by the equation $T = FL$, where "L" stands for lever arm.

Referring to the accompanying figure, suppose that mass "m" is attached to the end of a crank which in itself has no weight and is pivoted at "O." Suppose also that force "F" which acts upon the mass is a steady force so directed that it is always at right angles (tangential) to the crank. Under these conditions, as the mass moves, the crank pulls it toward the center and causes it to travel in a circular path. At the same time, the tangential force "F" causes it to move along the circle with increasing speed, giving it an acceleration along this path equal to "F/m." Thus,

$$a = \frac{F}{m}$$

Meanwhile, the crank will have an increasing angular velocity, and if the crank is of length r, its angular acceleration can be expressed as α (alpha) $= \frac{a}{r}$, where α is the instantaneous angular acceleration.

The angular acceleration of a body is defined as its time rate of change of angular velocity, or in formula: $\alpha = \frac{dw}{dt}$, where "dw" is the infinitesimal change of angular velocity occurring in the infinitesimal time interval "dt."

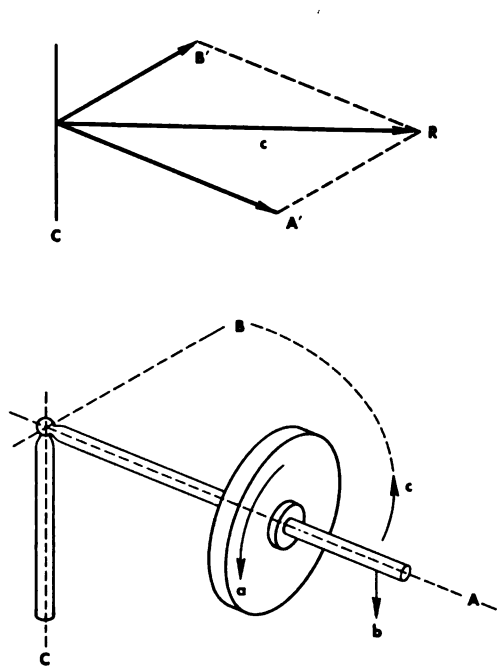
The force "F" acting on the mass "m" produces a torque "T" about the axis "o" which may be expressed by the equation $T = Fr$.

Precession

When two angular velocities about separate axes are added to a body, an angular motion about a third axis is produced. This new motion is called precession. This type of motion is illustrated below in the form of a wheel mounted loosely between collars on an axle "A." Assume that the wheel is rotating. The shaft does not rotate with the wheel, and one end of it is set horizontally upon a pivot on the vertical standard "C." The other end of the shaft is unsupported. If the wheel were not revolving, the free end of the shaft would drop and the entire assembly would fall off the vertical support.

But with the wheel revolving, the tendency for the free end of the shaft to drop causes the wheel and the shaft to describe horizontal circles about the pivot; in other words, the wheel precesses in the horizontal plane.

Let's now determine how precession develops. The angular velocity of the wheel rotating in the direction "a" about axis A is represented by vector "A" parallel to shaft "A." The gravitational pull of the earth upon



Precession of a revolving wheel about one end of its axle

the wheel produces a torque which tends to pull the entire assembly about axis "B" in the direction indicated by arrow "b".

This torque produces an angular acceleration and gives the body an additional angular velocity about axis "B" which is represented by the vector "B." If this velocity "B'" is added to the rotational velocity "A'" both being in the same horizontal plane, the resultant will be "R." Therefore, shaft "A" will shift its position to point in direction "R", turning about the vertical axis "C" as indicated by arrow "c." This motion of precession takes place about axis C which is perpendicular to both axis "A" and axis "B."

The precession continues, for as soon as the shaft reaches position "R," the assembly is subject to another torque due to the tendency of the free end to drop. The corresponding change in angular velocity is at right angles to "R" and a new resultant is formed to which the shaft next progresses, and so on. This progression of the shaft is not in discrete steps, but is actually a continuous procession of infinitesimal angular shifts, producing a uniform velocity of precession.

It can be shown that the angular velocity of precession, Ω (omega), of a wheel is equal to the Torque T, which tends to change the direction of the axis, divided by the product of the angular velocity ω of the wheel and its moment of inertia (I):

$$\Omega = \frac{T}{\omega I}$$

The above physical principles are embodied in the theory and operation of a gyroscope which finds a multitude of applications in missile guidance and control systems.

THE ACTION OF GYROSCOPES EXEMPLIFIES PRECESSION. A gyroscope exemplifies the physical principles and forces to a greater degree than any other single device. One form of gyroscope that you are familiar with is the earth. The earth resembles a huge gyroscope rotor spinning freely in space about an imaginary axis through its poles. Because of the rapidity of its rotation, the earth maintains its position in the universe and remains constantly in the same plane of rotation with its poles pointing approximately in a constant direction in space.

Aside from variations in physical form and certain variations resulting from bearing-friction and structural unbalance, the gyroscope represents the same physical phenomena in action as the earth, and the same forces are involved.

The gyroscope basically consists of a wheel or disk so mounted as to be able to spin rapidly about an axis, which can be called the axis of symmetry; and the wheel also is mounted so as to be free to rotate about one or both of two axes which are perpendicular to each other and to the axis of spin. Three types of gyroscopes are shown on the right.

If, when the wheel is spinning, a torque or twisting moment is applied about one of these two axes (axis of torque), the moment will produce precession about the other axis which then becomes the axis of precession. Under certain conditions, depending upon their orientation, these two axes may be referred to as the horizontal and vertical axes.

Another force which acts on a body to a small degree is Coriolis force, named for the French engineer G. G. Coriolis, who first brought it to attention.

Coriolis force is a deflecting force acting on a body in motion and occurring as a result of the earth's rotation. The force diverts horizontal motions to the right in the Northern Hemisphere and to the left in the Southern Hemisphere. This force is not present in stationary bodies.

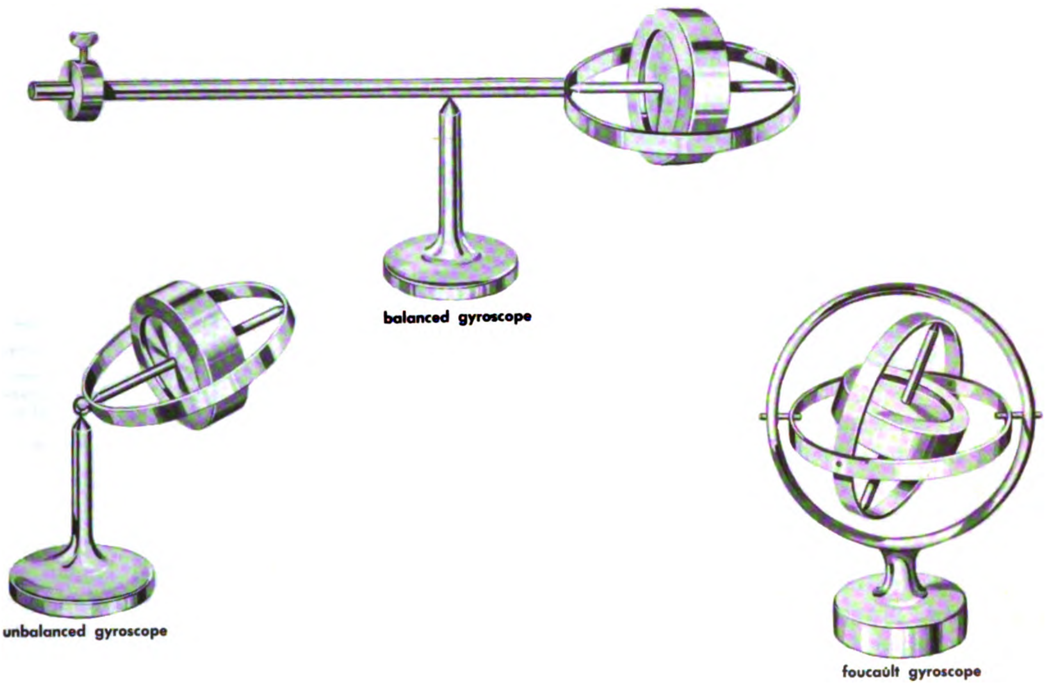
Gyroscopes are classified into three basic types: balanced, unbalanced, and Foucault. These types are shown on the next page.

A balanced gyroscope is one placed at one end of a horizontal rod and balanced on a pivot by means of a counterweight near the other end.

An unbalanced gyroscope is one in which the rod and counterweight beyond the pivot are omitted.

The Foucault gyroscope is one in which the spinning wheel is supported by gimbals so that it can turn freely about any axis, and the parts are so balanced that gravity does not exert a torque upon the wheel.

The spinning gyroscope offers considerable resistance, depending on the angular mo-



Basic types of gyroscopes

mentum, to any torque which would change the direction of the axis of spin; this property, known as *rigidity*, makes the gyroscope suitable for use as a stabilizer in such cases as resisting the roll of a ship or airplane, maintaining equilibrium of a moving body, and functioning as a steering device.

If the motion of a Foucault gyroscope about one of its axes is restrained, and it is given a rotational motion about another axis, the axis of spin will place itself parallel to the latter axis. This principle is exemplified in the gyrocompass.

Gyroscopes in many forms and modifications are used in the control and guidance systems of guided missiles. These will be treated at greater length and detail in connection with those subjects.

This section can be summarized by considering a gyroscopic element as a system with the following properties:

a. Is nonpendulous, and has freedom to rotate about a point.

b. Includes a symmetrical body spinning at a constant high rate about its axis of symmetry (axis of spin).

c. Has all angular momentum of the system effectively concentrated about the spin axis.

d. Is effectively a *viscous damper* with one end fixed in inertial space.

Inertial space may be conceived as a region in which all accelerations are assumed to be zero and all gravitational forces (mass attractions) are in a state of equilibrium.

With acceleration assumed to be zero, it follows that the unaccelerated space must also be nonrotating since rotation always involves acceleration.

Inertial space conforms to Newton's laws of motion, as exemplified in applied force being equal to time rate of change of linear momentum. Linear momentum is a vector quantity representing momentum of translation, and is equal to the product of the mass and velocity of the center of the mass.

In our concept of inertial space, which we generally associate with the "fixed" stars, the vector sum of all forces acting throughout the region is equal to zero.

Energy

A body possesses *energy* if it is capable of doing work.

When work is performed by a body, its energy is reduced by an amount exactly equal to the work done. If a body is set in motion, it will exert a force and consume energy in coming to rest. Thus it has done work.

A moving body always possesses energy by virtue of its motion: such energy is called *kinetic energy*. A body at rest may possess *potential energy* by virtue of its position. A body possesses such energy when it is at a position from which it is capable of doing more work.

The law of *conservation of matter* states that matter can neither be created nor destroyed. Accordingly, the total amount of matter in the universe remains constant.

From the two laws of physics presented above, it can be deduced that when work is performed energy is released or set in motion. Energy is not created. In the same process, matter may be changed in form; it is not destroyed.

Another form of mechanical work occurs when one body moves over another. Work is done against friction which exists between the bodies. Thus, when mechanical work is done on a body, it can be entirely accounted for by one or more of the following effects:

a. Increase in the kinetic energy of the body.

b. Increase in its potential energy.

c. Production of heat due to friction.

Equilibrium

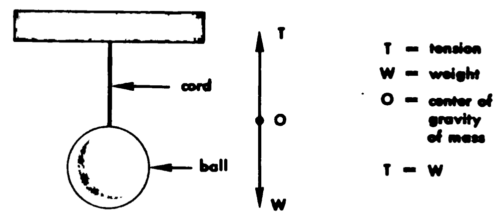
A body which continues in a state of rest is said to be in *static equilibrium*. The simplest condition of equilibrium is that in which a body is acted upon by only two forces. In such a case, one of these forces must be equal and opposite to the other, as indicated in the illustration above right.

The first condition of equilibrium requires that the vector sum of all the forces acting upon a body along any direction must equal zero.

The second condition of equilibrium requires that the torques acting upon a body shall be balanced; the clockwise torques must be equal to the counterclockwise torques. Notice in the diagram on the right that the algebraic sum of torques equals 0. In the diagram:

$$\frac{m_1 l_1 = m_2 l_2}{10 \times 1 = 5 \times 2}$$

An important theorem, which follows, is based upon the second condition of equilibrium. "If three nonparallel forces acting upon a body produce equilibrium, their lines of action must pass through a common point."



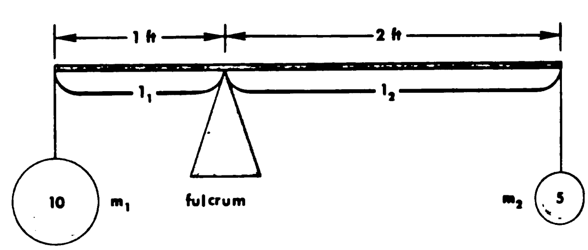
Simple equilibrium

This theorem can be demonstrated by taking moments of forces about the intersection of two of the lines of action and showing that the lever arm of the remaining force must be zero. Suppose that a body is in a state of equilibrium under the action of three forces, as shown incorrectly in the above right diagram.

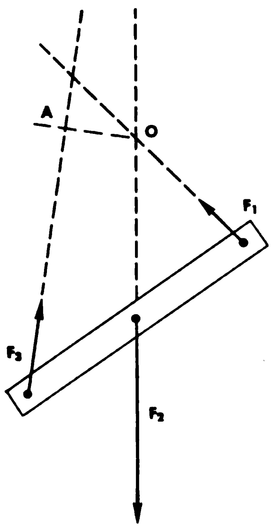
Since their lines of action are not parallel, any two, if extended, will meet at some point. Thus, the lines of action of "F_1" and "F_2" meet at "O." From this point draw a line "OA" perpendicular to the third force "F_3," meeting its line of action at "A."

Apply the second condition of equilibrium (algebraic sum of torques = O), by taking the moments of all forces about "O," and setting their sum equal to zero. The moments of "F_1" and "F_2" are both zero, since the lines of action of these forces pass through "O," and thus their lever arms are zero.

Since the body is in equilibrium, we would expect the moment of the remaining force



Fulcrum-and-lever example of balanced torques

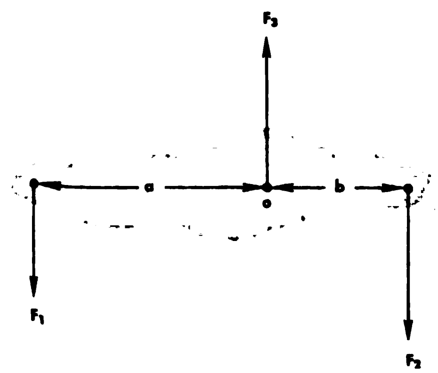


Body under the action of three forces

"F_3" about "O" to be zero; and since the force "F_3" is not zero, its lever arm "OA" must be zero instead of as shown. That is, the line of action of "F_3" must pass through O. Thus the lines of action of three forces in equilibrium pass through a common point.

You may deduce from the foregoing example that when a body is in a state of equilibrium under the action of three forces, the resultant of any two of these forces is equal and opposite to the third force.

The same principle employed in the preceding example may be used in determining the resultant of two parallel forces.



A body in equilibrium under the action of three parallel forces

Suppose that the body shown in the diagram to the left below is in equilibrium under the action of three parallel forces, F_1, F_2, and F_3. From the first condition of equilibrium, $F_3 = F_1 - F_2$. The resultant of the forces "F_1" and "F_2" is equal and opposite to "F_3." From the second condition of equilibrium $F_3b - F_1a = O$, when the moments of force are taken about point "O." This equation shows an important relation between the forces and their lever arms:

$$\frac{F_2}{F_1} = \frac{a}{b}$$

Therefore, the resultant of two parallel forces has the same direction as the forces and is equal to their sum, and the resultant's line of action divides the distance between them into two parts which are *inversely proportional* to the respective forces.

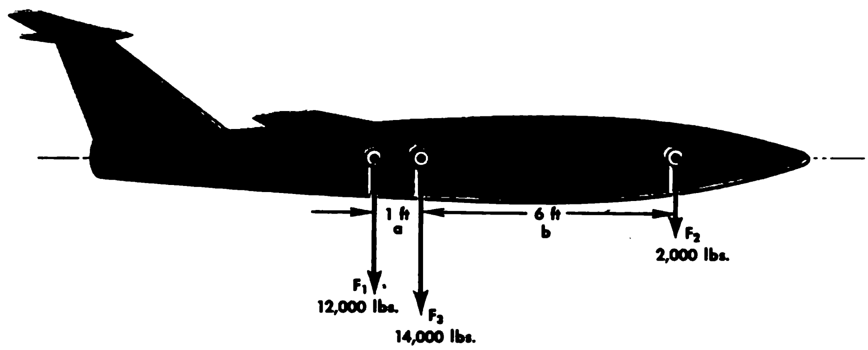
The resultant of any number of parallel forces may be found by first determining the force which, included with the forces given, will produce equilibrium. This force is called the *equilibrant* of the given forces, and *the resultant is equal and opposite to the equilibrant and acts along the same line.*

The attraction which the earth exerts upon a body extends to every particle of matter of which the body is composed. Thus, the weight of a body may be considered as an aggregation of parallel forces acting upon the individual particles in the body. If these parallel forces are replaced with their resultant, this single force is exactly equivalent to them.

For any given object there is a particular point through which the resultant of the weight forces will always pass regardless of the orientation of the object. This point is called the CENTER OF GRAVITY (C. G.) of the body. Hence, the weight of an object, although actually a system of parallel forces acting upon all its component parts, can be correctly represented by a single force acting downward at the center of gravity.

There are instances when it is desirable to find the center of gravity of a missile. For example, the principles stated above are applicable when the center of gravity of a missile has shifted due to adding weight to the craft.

In the following sketch, if the weight of the



Change of center of gravity resulting from added weight to a missile

missile, less the warhead, is 12,000 pounds and its center of gravity is at point "A", the addition of the 2,000-pound warhead with a center of gravity at point "B" will produce a resultant downward force of 14,000 pounds at a new center of gravity as indicated at point "C." If the distance between points "A" and "B" is seven feet, the new center of gravity will be one foot (a) forward of the original C. G. as shown at point "A." The one-foot change of position of the C. G. is found in the following manner:

$F_2/F_1 = a/b, \; F_1 \times a = F_2 \times b$

$12{,}000 \times a = 2{,}000\ (7-a) \quad a+b=7$ feet

$12{,}000\ a = 14{,}000 - 2{,}000\ a$

$14{,}000\ a = 14{,}000$ or $a = 1$

Harmonic Motion

The principles underlying *harmonic motion* are closely associated with those of *circular motion*. These principles pertain to both mechanical and electrical systems.

Harmonic motion is defined as a vibratory motion in which the acceleration of the vibrating body and the *restoring force* acting upon it are proportional to its displacement from the midpoint of its path and are directed toward that point.

The relationship between harmonic motion and circular motion is exemplified in the sketches above right. If the pin on the rim of the rotating disk is viewed along the plane of the disk, it appears in the mirror to be vibrating *to and fro* along a path equal in length to the diameter of the disk. If the disk rotates at constant speed, the projection cast by the uniform circular motion in the plane of the disk is harmonic motion.

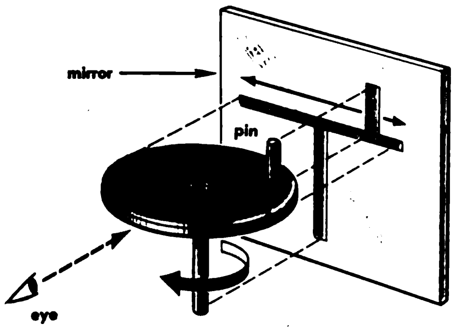
Similarly in the example of the engine, you will note that the circular motion of the wheel is applied to a piston which is so mounted that its motion is restricted to a single plane. In this case, when the wheel is rotated at constant speed, the drive-pin travels in a circle at a uniform rate but the piston receives only horizontal projection of this motion. Thus, it slides to and fro in the cylinder through a distance equal to the diameter of the circle described by the pin. The to and fro movement of the piston is harmonic motion.

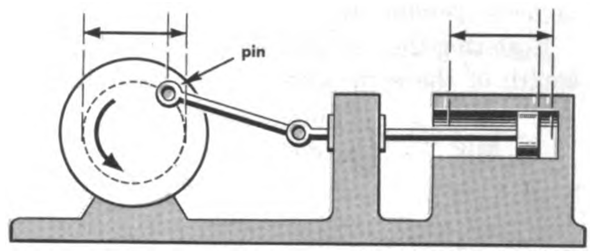
You must bear in mind that not all to-and-fro motion is harmonic motion. Many machines employ parts which move back and forth repeatedly along the same path in equal time intervals but do not conform to the definition of harmonic motion. Such motion may be classed as *periodic motion* since it recurs in equal periods of time. But such motion is not harmonic since the rate of change of the motion is not linear. Thus, it does *not* necessarily follow that a body which travels in a circular motion describes harmonic motion.

Pendulums

As you study further into the make-up of a missile, you will come to realize that many basic units used in missile control and guidance systems utilize the principle of the *pendulum*. Examples of such units are inertial switches, stabilized platforms, timing standards, and compensating devices.

A pendulum is defined as a body so suspended from a fixed point as to swing freely to and fro under the combined action of gravity and momentum.

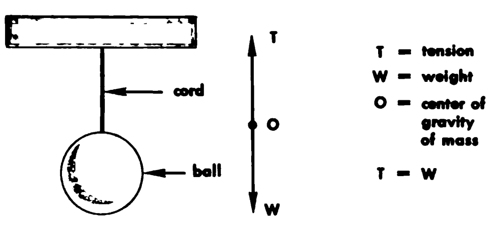




Harmonic motion compared to circular motion

Usually, the theory of the pendulum is explained in terms of a *simple* or *mathematical pendulum*, which is shown below. The simple pendulum is a purely theoretical device consisting of a particle or material point suspended by a thread without weight and oscillating without friction. By visualizing the pendulum as existing under these conditions, the time of vibration can be determined by considering the pendulum to be displaced slightly and then by studying the forces acting upon it. An analysis will show that the unbalanced force is proportional to the displacement and opposite in direction. The vibration (oscillation) of the pendulum is harmonic, and the equations of harmonic motion can be used to determine the period of vibration.

$$T = 2\pi \sqrt{\frac{l}{g}} \qquad T = \text{Period}$$



Simple, or mathematical, pendulum

The time of vibration of a pendulum is determined from the amplitude of vibration (or the acceleration due to gravity (g) at the point where the pendulum is located) and its length (1) (or the distance between the axis of suspension and the axis of oscillation).

By using a pendulum of known length and measuring its period of vibration, the value of gravity can be determined. The equation

$$T = 2\pi \sqrt{\frac{l}{g}}$$

applies only to small values of displacement angle; the values need not be constant. As the vibration of the pendulum continues, the amplitude of oscillation continually decreases, but the period remains practically unchanged. It is this property of the pendulum which makes it especially suitable for controlling the escapement mechanisms of timekeeping devices.

Another type of pendulum, the *physical* or *compound* pendulum, is a body which vibrates in the manner of a pendulum but in which the mass is distributed and not concentrated as in the simple pendulum.

The *center of oscillation* of a physical pendulum is that point at which the concentration of the whole mass of the pendulum would produce no change in its period of vibration. If the mass were so concentrated, the physical pendulum would be identical to a simple pendulum having the same period of vibration.

$2\pi \sqrt{\frac{l}{g}}$ represents the period of the simple pendulum.

$2\pi\sqrt{\frac{1}{Mgh}}$ represents the period of the physical pendulum.

Equating the two preceding expressions, the length of the equivalent simple pendulum is:

$1=\frac{I}{Mh}$, where I = moment of inertia; h = distance between axis of suspension and the center of gravity.

The center of oscillation of a physical pendulum is separated from the axis of suspension by the distance $\frac{I}{Mn}$; this expression also shows the length of the simple pendulum which has the same period of vibration as the physical pendulum. The center of oscillation may be interchanged with the center of suspension without affecting the period. These characteristics of the pendulum lend themselves to the measurement of vibration periods, centers of gravity, and other factors of missile configuration and their probable effects upon the flight characteristics of the missile.

The center of oscillation is also the *center of percussion* because at this point the pendulum can be struck with no resulting jar on the axis of suspension. This property is of value in the determination of correct suspension of components for absorbing shock. When struck at any point other than the center of percussion, the pendulum not only tends to rotate but the axis of suspension force receives a motion of translation.

Still another type pendulum, a *torsion pendulum*, is a suspended body that vibrates not be swinging, but by rotating with an alternate twisting and untwisting movement.

The characteristics of this type pendulum make it especially suitable for use in timing devices, balance wheels of watches, torque-measuring instruments, torque-compensation devices, magnetometers, strain-gauges, and others.

You can consider a torsion pendulum as a weight attached to a vertically suspended wire. When the wire is twisted and released, it describes a series of angular vibrations similar to the linear vibrations of harmonic motion.

When the suspension wire is twisted, an oppositely directed restoring torque which is proportional to the angular displacement is set up within the wire. Just as in linear harmonic motion, a restoring force is set up which is proportional to the linear displacement. The two types of motion are closely analogous, and angular harmonic motion can be represented most simply by comparing it directly to linear harmonic motion.

You can take the equation

$$\frac{F}{x}=4\pi^2n^2m$$

which expresses linear harmonic motion, and replace each linear quantity with its corresponding angular quantity and derive a corresponding expression for angular harmonic motion:

$$\frac{T}{\phi}=4\pi^2n^2I$$

where "T" represents the restoring torque, "ϕ" is the corresponding angular displacement, "n" is the frequency, and "I" is the moment of inertia of the vibrating body about the axis of rotation.

The ratio of the torque to the corresponding twist is a constant determined by the stiffness of the suspension wire and is essentially negative (opposite to the applied force). You can replace the ratio $\frac{T}{\phi}$ with τ (tau) and rearrange the expression. The new arrangment shows the frequency of angular vibration to be:

$$n=\frac{1}{2\pi}\sqrt{\frac{\tau}{I}}$$

Consequently, the period is expressed by

$$T=2\pi\sqrt{\frac{I}{\tau}}$$

which is analogous to

$$T=2\pi\sqrt{\frac{l}{g}}$$

This latter equation expresses the period of linear vibration.

The torsion principle can be used to determine the moment of inertia of a body by supporting the body as a torsion pendulum and measuring its period and the angle of twist produced by a measured torque.

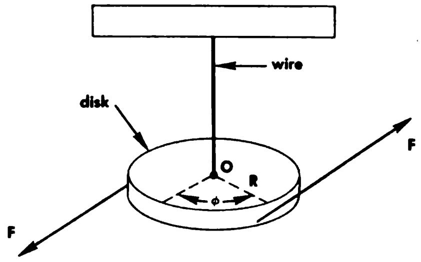
By transposing equation

$$T=2\pi\sqrt{\frac{I}{\tau}}$$

the moment of inertia "I" of the suspended body can be determined:

$$I = \sqrt{\frac{\tau T^2}{4\pi^2}}$$

In the accompanying figure, the forces "F" acting at the distance "R" from the center "O" produce a torque "2RF."



Torsion pendulum

MECHANICS OF GASES

Gas is one of the three basic forms in which matter exists, the other two forms being liquid and solid.

Gas can be defined as an aeriform fluid which has neither independent shape nor volume but tends to expand in an unlimited manner. This tendency to expand is due to the fact that the molecules of a gas are relatively far apart and are continuously in motion throughout the entire space occupied by the gas.

In many respects, gases resemble liquids. Since both are capable of flowing, they are commonly referred to as fluids.

The two major differences between gases and liquids are in respect to compressibility and expansion. Gases are highly compressible, while most liquids are but slightly so. Gases tend to completely fill any closed vessel in which they may be contained, while most liquids, like solids, fill a container only to the extent of their normal volume.

Both gases and liquids exert pressure upon the surfaces with which they are in contact, and each exerts an upward buoyant force that is in accordance with the principle of Archimedes, which states that "a body submerged wholly or partially in a fluid is buoyed up by a force equal to the weight of the fluid which it displaces."

Flowing gases tend to conform to Bernoulli's theorem on liquids when their compressibility is considered as a factor. The velocity of their effusion through an orifice may be calculated in the same manner as for liquids. Bernoulli's theorem states that if no work is done on or by an incompressible liquid as it flows, the total head remains unchanged. "Total head" refers to the total energy of a moving liquid at a given point. From this theorem, it can be deduced that when a compressible fluid flows, the total head varies in proportion to the degree of compressibility.

The mathematics of Bernoulli's theorem are simple when applied to hypothetical fluids considered as being incompressible or to water which is but very slightly compressible. But the mathematics become complex when applied to gases which are highly compressible in varying degrees and whose densities differ with differences in pressure.

However, the general effect applies to all fluids. In respect to gases, it is enough to keep in mind that as a flowing stream of gas increases in velocity, its pressure decreases, and vice versa. This effect is noticeable in the case of lift exerted upon a moving airfoil and governs the design of the airfoil for specific applications. Lift is discussed in Chapter 2.

Gases, like liquids, adapt themselves to the shape of the vessel in which they may be contained and, having no elasticity of shape, are unable to exert shearing stresses other than those due to their viscosity.

A gas which can be liquefied by pressure alone is termed a vapor. At room temperatures, steam and carbon dioxide are called vapors; but air, hydrogen, and nitrogen are called gases.

Kinetic Theory of Gases

The simple structure of gases makes them readily adaptable to mathematical analysis from which has evolved a detailed theory of the behavior of gases. This has been called the *kinetic theory of gases*. The theory assumes that a body of gas is composed of identical molecules which behave like minute elastic spheres, spaced relatively far apart and are continuously in motion.

The degree of molecular motion is dependent upon the temperature of the gas. Since the molecules are continuously striking against each other and against the walls of the container, an increase in temperature with the resulting increase in molecular motion causes a corresponding increase in the number of collisions between the molecules. The increased number of collisions results in an increase in pressure because a greater number of molecules strike against the walls of the container in a given unit of time.

If the container were an open vessel, the gas would tend to expand and overflow from the container. However, if the container is sealed and possesses elasticity (such as a rubber balloon), the increased pressure will cause the container to expand.

You may have noticed that when making a long drive on a hot day, the pressure in the tires of your automobile increases and that a tire which appeared to be somewhat "soft" in cool morning temperature may appear normal at a higher midday temperature.

Such phenomena as these have been explained and set forth in the form of laws pertaining to gases and tend to support the *kinetic theory*.

At any given instant, some molecules of a gas are moving in one direction, some in another direction; some are traveling fast while some are traveling slowly; some may even be in a state of rest. The combined effect of these varying velocities corresponds to the temperature of the gas. In any considerable amount of gas, there are so many molecules present that in accordance with the "laws of probability," some average velocity can be found which, if it were possessed by every molecule in the gas, would produce the same effect at a given temperature as the total of the many varying velocities.

You know that it requires energy to raise the temperature of a substance, so you can assume that the temperature of a gas is directly proportional to the mean kinetic energy of the gas molecules. You must realize that this mean velocity must be such that it imparts the same kinetic energy to a given number of molecules of specific mass as would be imparted by their various individual velocities.

Atmospheric Pressure

A gas that is vital to missile flight is air. The mass of air surrounding the earth and held to it by gravitational attraction exerts a pressure upon the earth's surface:

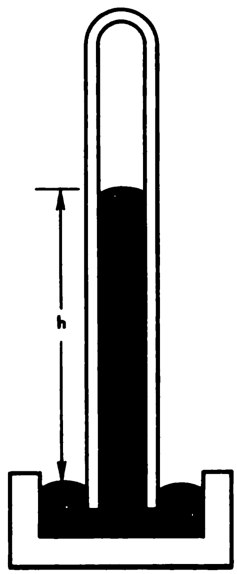
$$p = hdg$$

where: h equals the height of the air layer
d equals the density of the air layer
g equals the gravitational attraction of the earth

This atmospheric pressure can be measured by any of several methods. The common laboratory method employes the mercury column (barometer). A mercury column consists of a glass tube approximately 34 inches in length, sealed at one end, then completely filled with mercury, and inverted in an open cup partially filled with mercury, as shown in the accompanying figure.

The mercury in the tube settles down, leaving an evacuated space in the upper end of the tube. The height (h) of the mercury column serves as an indicator of atmospheric pressure.

At sea level and at a temperature of 0° C, the height of the mercury column is 76 cm or approximately 30 inches, which represents a pressure of 14.7 pounds per square inch. The 30-inch column is used as a reference standard.



Mercury column for measuring atmosphere pressure

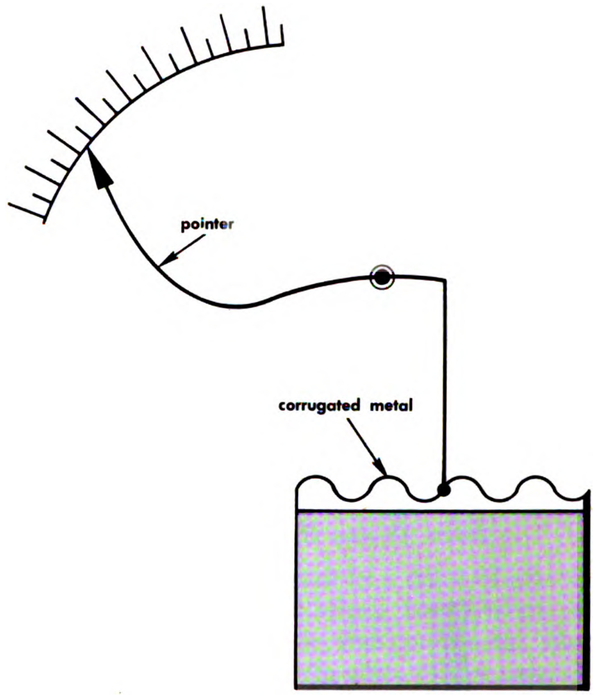
At higher levels, the atmospheric pressure on the surface of the mercury in the open cup is less than at sea level; hence the column of mercury in the tube settles lower. These variations in the height of the mercury column represent changes in atmospheric pressure which may be calibrated in terms of altitude with reference to sea level.

Another device used to measure atmospheric pressure is the *aneroid barometer*, pictured at the right. This barometer consists of a small sealed metal drum that has been partially evacuated and which has one side or end made of thin corrugated metal. This thin metal moves in or out with the variation of pressure on its external surface. This movement is transmitted through a system of levers to a pointer. The pointer is pivoted so that it sweeps across a graduated scale calibrated either in pounds per square inch or in feet to designate an altitude corresponding to the indicated pressure. This device is used in aircraft altimeters and as an end instrument or transducer in telemetering circuits to obtain information pertaining to pressures and altitude.

The atmospheric pressure does not vary uniformly with altitude. It changes more rapidly at lower altitudes because of the compressibility of the air which causes the air strata close to the earth's surface to be compressed by the air masses above them. This effect, however, is partially counteracted by contraction of the upper strata due to cooling. The cooling tends to increase the density.

Altitude can be approximately estimated from a knowledge of the corresponding atmospheric pressure. When proper compensation is made for humidity, density, temperature, and any other variables, fairly accurate results can be obtained from the barometric altimeter.

Atmospheric pressures are quite large, but in most instances practically the same pressure is present on all sides of objects so that no single surface is subjected to a great load. To exemplify the tremendous atmospheric pressure which may exist on a surface, consider the pressure present on the face of an ordinary 5-inch cathode-ray tube as used in oscillographs. At sea level the pressure on the face of this tube is approximately 300 pounds. Consider also an average 17-inch rectangular television tube. It would have a pressure of almost 2400 pounds on its screen surface at sea level.



Aneroid barometer

The force produced by atmospheric pressure is calculated by the use of the equation

$$F = pA,$$

where: F = total pressure on a surface

p = atmospheric pressure per square inch under specific conditions of altitude and density

A = the area of the surface in square inches

Compressibility of Gases

Compressibility is an outstanding characteristic of gases. The simple relationship between the pressure of a gas and its volume is stated in Boyle's law, which states that "the volume of a confined body of gas varies inversely as the absolute pressure, provided that the temperature remains constant."

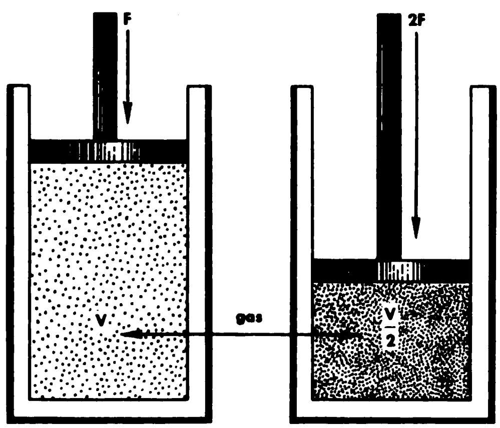
This law can be demonstrated by confining a quantity of gas in a cylinder which has a tightly fitted piston. A force is then applied to the piston so as to compress the gas in the cylinder to some specific volume. When the

force applied to the piston is doubled, the gas is compressed to one-half its original volume, as indicated below.

If "p_A" and "V_A" represent the pressure and volume under one set of conditions, and "p_B" and "V_B" represent the pressure and volume under another set of conditions, then:

$$V_A : V_B = p_B : p_A$$
$$p_A V_A = p_B V_B$$

A gas which conforms to Boyle's law is termed an ideal gas. When pressure is increased upon such a gas, its volume decreases proportionally and its density is increased. So it follows that the density of a gas varies directly as the pressure, if temperature remains constant as in the case of an ideal gas. Density also varies with temperature, since gases expand when heated and contract when cooled.



Gas compressed to half its original volume by a double force

The specific gravity of a gas is computed with reference to air as a standard. Therefore, the density of air has been determined accurately. One liter (1000 cubic centimeters) of air at 0° Centigrade and 76 cm of mercury weighs 1.293 grams. Thus, the density of air is 0.081 pounds per cubic foot under standard conditions.

By dividing the density of a gas by the density of air, the relative specific gravity of the gas is found. The following table shows the density and specific gravity of some common gases:

GAS	DENSITY gm/liter	DENSITY lb/cu ft	SPECIFIC GRAVITY
Air 0° C 76 cm Hg	1.293	0.081	1.000
Air 20° C	1.205	0.0755	0.932
Steam 100° C	0.598	0.037	0.462
Hydrogen	0.090	0.0056	0.069
Helium	0.179	0.011	0.138
Nitrigen	1.251	0.078	0.967
Oxygen	1.429	0.089	1.105
Carbon dioxide	1.977	0.123	1.529

Dalton's Law

If a mixture of two or more gases which do not combine chemically is placed in a container, each gas expands throughout the total space and the absolute pressure of each gas is reduced to a lower value called its *partial pressure*. This reduction is in accordance with Boyle's law. The pressure of the mixed gases is equal to the sum of the partial pressures. This fact was discovered by Dalton, an English physicist, and is set forth in Dalton's law which states that "a mixture of several gases which do not react chemically exerts a pressure equal to the sum of the pressures which the several gases would exert separately if each were allowed to occupy the entire space alone at the given temperature and in the same volume as each is present in the mixture."

AVOGADRO'S LAW. An Italian physicist, Avogadro, conceived the theory that "at the same temperature and pressure, equal volumes of different gases contain equal numbers of molecules." This theory was proven by experiment and found to agree with kinetic theory, so it has come to be known as *Avogadro's law*.

Viscosity

Another property common to both gases and liquids is *viscosity*, which is defined as *resistance to flow*. Viscosity is due to molecular

friction and is less pronounced in gases than in liquids due to the greater space between molecules in the gases.

However, the viscosity of gases noticeably retards their flow through pipes or tubes and also retards the passage of other bodies through a stationary body of gas. Increasing the temperature of a gas increases its viscosity as a result of the increase in the movement of its molecules.

Viscosity can be considered as the resultant of *cohesion* and *adhesion*. Cohesion is the molecular attraction which tends to unite all particles of a body throughout its mass. Adhesion is the molecular attraction exerted between the surfaces of bodies which are in contact. Airfoils are streamlined to reduce the effect of this air friction and, thus, increase the velocity of the aircraft of which they may be components.

All of the phenomena of gases mentioned in this section and the laws to which they conform must be considered in the design of aircraft from the viewpoints of both airframes and propellants. Specific applications of the mechanics of gases are treated in detail in other chapters of this manual in connection with subjects to which they are pertinent.

Physics Thus Far

Let's quickly review what has been covered in this chapter so far.

Keep in mind that inertia is a property of matter by which matter tends to remain at rest or, if in motion, tends to remain in motion in the same direction unless acted upon in either case by some external force.

Velocity is a representation of speed and a specific direction of motion. Angular velocity refers to a variation in direction of a moving object. Acceleration is the time rate of change of velocity. It represents motion in which the velocity changes from point to point.

Centrifugal and centripetal forces can be illustrated by a ball attached to a cord. When the ball is whirled at the end of the cord, the ball pulls outwardly, exerting a centrifugal force on the cord. The centrifugal force causes the cord to become taut. The cord then exerts an inward pull, which is centripetal force.

Angular acceleration is produced by torque. Torque is a rotational effect exerted on a body and is measured by the product of the force acting on the body and the perpendicular distance from the axis of rotation to the line of action of the force.

When two angular velocities about separate axes are added to a body, an angular motion about a third axis is produced. This new motion is called precession. The operation of a gyroscope exemplifies the phenomena of precession.

Moving bodies possess kinetic energy by virtue of their motion. A body may possess potential energy by virtue of its position; a body possesses potential energy when it is at a position from which it can do work.

A body which continues in a state of rest is said to be in static equilibrium. Equilibrium of a body is dependent upon two conditions. The first condition requires that the vector sum of all the forces acting upon a body along any direction must equal zero. The second condition of equilibrium requires that the torques acting upon a body shall be balanced.

Harmonic motion is defined as a vibratory motion in which the acceleration of the vibrating body and the restoring force acting upon it are proportional to its displacement from the midpoint of its path and are directed toward that point. The to-and-fro motion of a pendulum describes harmonic motion.

Remember that gases are highly compressible and that they tend to completely fill any closed vessel in which they may be contained. Also remember that as a flowing stream of gas increases in velocity, its pressure decreases, and vice versa. Temperature affects the pressure that a gas exerts against the walls of a confining container. An increase in temperature results in an increase of molecular motion which in turn causes an increase in pressure. This behavior is referred to as the kinetic theory of gases.

A phase of physics which is important in the missile field is optics or, in other words, the science of light. Optics are discussed in the following section.

Optics Involved in Guided Missile Operation

Some of the systems of navigation and homing used in modern missiles are based upon the behavior and phenomena of light or some form of radiant energy possessing characteristics similar to those of light. Examples of such systems are the *heat-seeking* and *light-seeking* homing systems and automatic celestial navigation.

To utilize the properties of light and radiant energy in missile systems, it is necessary to transform or adapt the radiant energies to a form of energy suitable for actuating or controlling the mechanical or electrical components employed in guiding and controlling missiles. This adaptation generally involves the use of an *optical system* in conjunction with the electrical and/or mechanical units used to control the flight of a missile or to determine its position. Thus, your primary concern is the method by which the sources of light or radiant energy are employed as reference points in homing and navigation systems.

To facilitate your understanding of the theory and operation of homing and navigation systems, some knowledge of the properties of light and of radiant energy is essential.

Modern theory considers light, both visible and invisible, as consisting of *quanta* (bundles) of energy which move as if guided by waves. The statistical behavior of the quanta is dependent upon the assumption that the energy of the quanta at any point is on the average equal to the intensity of the wave system at that point.

This *wave theory* of light assumes that light is transmitted from luminous (light-emitting) bodies to the eye and other objects by an undulatory or vibrational movement. The velocity of this transmission is approximately 186,300 miles per second, and the vibrations of *ether* (conducting medium of light in space) are transverse to the direction of propagation of the wave motion. These waves vary in length from approximately 3.85 to 7.60 ten-thousandths of a millimeter.

The impression of *color* produced when the light energy impinges on the retina of the eye varies in a complex way with the wavelength, the amplitude of vibration, and various other factors and conditions, some of which are beyond the scope of this manual.

Waves of similar character, but whose lengths are above or below the limits mentioned in the preceding paragraph, are not perceptible to the average eye under normal conditions. The very short waves between 1.0 and 3.85 ten-thousandths of a millimeter in length constitute *ultra-violet* light and are made known by their photographic or other chemical action. Those waves which are longer than 7.60 ten-thousandths of a millimeter are the *infrared* waves and are detected by their thermal (heat) effects.

The *electromagnetic theory of light* as set forth by Maxwell, English physicist, holds that these waves, including those of light proper, are the same kind as those by which electromagnetic oscillations are propagated through

ether and that light is an electromagnetic phenomenon.

The most important phenomena of light are reflection, refraction, dispersion, interference, and polarization. It is one or more of these phenomena of light which acts through a suitable optical system as a medium for missile homing and navigation, so it is with them that we shall concern ourselves in this chapter.

Before undertaking a study of optics, let's define optics. Optics is the science pertaining to light, light's origin and propagation, the effects to which it is subject and which it produces, and other phenomena closely associated with it. *Geometrical optics* is concerned with the optical phenomena associated with reflection and ordinary refraction, but only insofar as they can be deduced mathematically from the simple laws of reflection which have been derived from observation and experimentation. *Physical optics* is concerned with the description and explanation of all optical phenomena in terms of physical theories, as wave theory in general, electromagnetic phenomena, quantum mechanics, and other light properties.

One of the mysteries pertaining to radiant energy such as light, heat, and electromagnetic waves is the medium through which it is conducted. Only because it absorbs the energy and changes it to some other form are we able to recognize its existence and determine its characteristics. As indicated earlier, the word *ether* is used to name the medium through which radiant energy is conducted. But it is not known what the medium actually is.

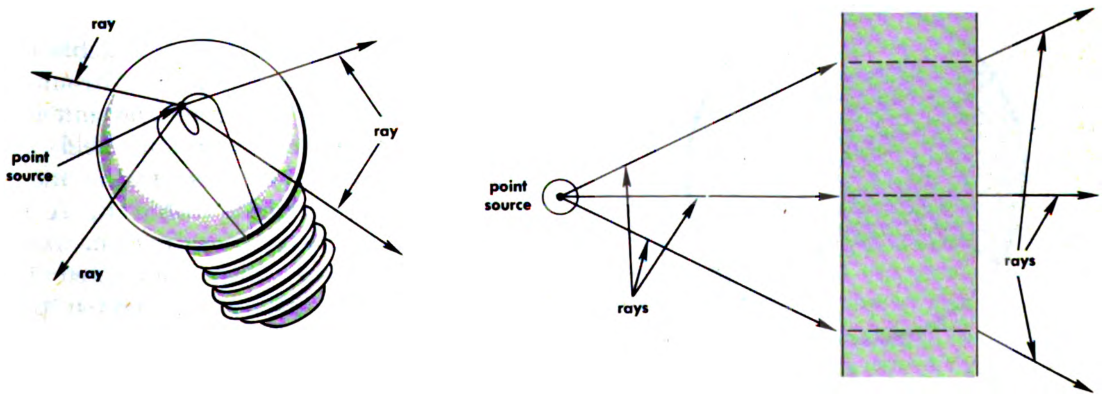
The most obvious fundamental property of light is that it travels in straight lines when passing through a homogeneous (uniform in density and composition) medium. The rectilinear (straight-line) propagation of light supports the idea that a ray of light is the rectilinear path in a homogeneous medium along which light is propagated or transmitted.

Thus, by choosing one point on a luminous body and from that point drawing a straight line in the direction of the propagation of light, you can represent a *ray* of light. From this point-source of light, you can draw an infinite number of rays, as shown below. This collection of rays, or cone of light, is referred to as a *pencil of rays*. Two rays from such a group are enough to locate the point-source of the light by simple geometric means. The point-source is the point of intersection of lines extended along the paths of the rays.

UNITS OF LIGHT INTENSITY

Luminous intensity or brightness of light represents the degree to which visible light is present in the radiant energy emitted by the source.

The retina in the human eye is sensitive to a relatively small portion of the radiant energy emitted by an incandescent body; therefore, to measure the relative intensities of visible light, we must employ standards and techniques specifically pertaining to visible light. Such standards and techniques constitute the science of *photometry*.



Light rays from a point-source

Light can be considered as a flow of radiant energy or *luminous flux* (expressed in ergs per second). Because of the variations of sensitivity of the human eye to different colors (different wavelengths of light), luminous flux cannot be measured visually in ergs per second. In place of the erg, a unit called the *lumen* must be employed for this purpose.

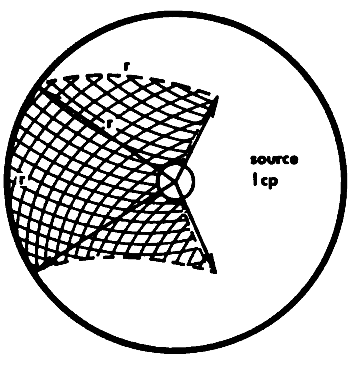
A lumen is the amount of light flux radiating from a uniform one-candle-power (1-cp) source throughout a solid angle of such size as to surround a unit area at a unit distance from the source.

Light flux refers to the rate at which a source emits light energy, evaluated in terms of its visual effect.

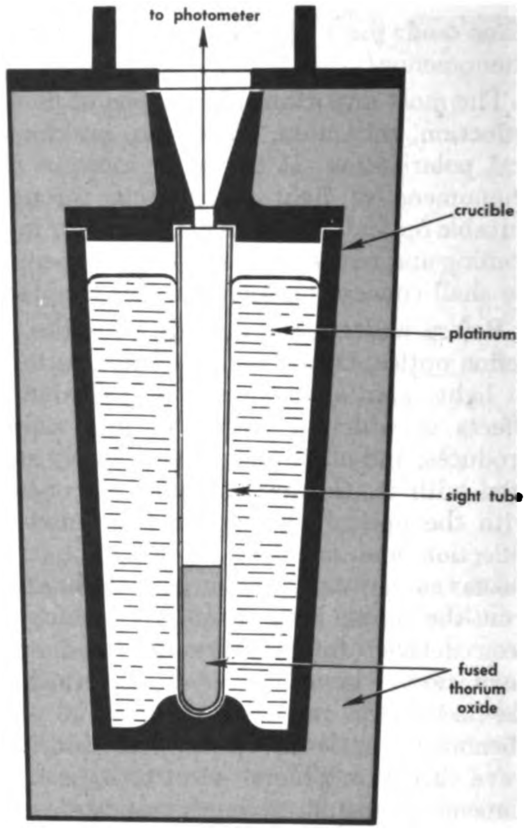
If you imagine the one-candle-power source to be located at the geometric center of a hollow sphere having a radius of one foot, then each square foot of the interior surface of the sphere receives one lumen of light. Since the total area of the sphere is four pi (4π) square feet, the total light emitted by the one-candle-power source is 4π lumens. The figure below of a sphere and solid angle illustrates this example.

Light from a one-candle-power source falling upon a unit area at unit distance represents one lumen.

Most sources of light have different luminous intensities along different directions. The average of the candle powers measured in all directions about a source of light is called the *mean spherical candle power*. Therefore, if a source having a mean spherical candle power of 1 cp emits 4π lumens of light flux, the total



Light falling on area r^2 at a distance "r" from 1-cp source equals 1 Lumen



National Bureau of Standards' Primary Standard of Luminous Intensity

flux (F) in lumens emitted by a source of mean spherical candle power (I_o) may be expressed by the equation: $F = 4\pi\ I_o$.

The primary standard of luminous intensity, developed by the National Bureau of Standards, consists of a glowing enclosure operated at the temperature of solidifying platinum (2046° kelvin or 2187° centigrade) and arranged as shown in the figure above. The platinum is contained in a crucible of fused thorium oxide or quartz, surrounded by a heat-insulating material. The unit is placed in an alternating magnetic field so that the platinum in the crucible is melted by the currents induced within it. A viewing tube of fused quartz or thorium oxide, and containing some finely powdered thorium oxide, is enclosed by the molten platinum and serves as a "black body" radiator. The brightness within this tube is considered to be 60 candle power per square centimeter when the metal, in cooling slowly, attains its

solidifying temperature. The new *standard candle* is therefore one-sixtieth of the luminous intensity of one square centimeter of a hollow enclosure at the temperature of solidifying platinum. This intensity is slightly less than that of the old *international candle*.

The amount of light flux which falls upon a surface and the area of the surface together determine the degree of illumination. The illumination is uniform only when a single source of light is employed and when all points on the illuminated surface are equidistant from the source.

A light intensity of one lumen per square foot is a *foot candle* and is the amount of illumination received on a surface one foot distant from a standard candle. The illumination of a surface is measured by the number of lumens incident upon a unit of area.

$$E = \frac{I}{d^2}$$

where "E" represents the total illumination, "I" equals the intensity of the source in candle power, and "d" equals the distance from source to surface.

If you consider a surface of area "A" as receiving a total light flux* "F", you can express the illumination of the surface in terms of lumens per unit of area, such as lumens per square foot. Expressed as an equation:

$$E = \frac{F}{A}$$

The degree of illumination which a light source produces upon a given surface depends upon the intensity of the source and its distance from the surface, provided that the rays of light pass through a uniform medium and strike the surface normally.

Increasing the intensity of the source produces a proportional increase in the light flux falling upon the surface. Increasing the distance of the surface from the source decreases the illumination of the surface by an amount proportional to the square of the distance. That is, doubling the distance between the source of light and the illuminated surface will reduce the illumination of the surface to one-fourth its original value.

This effect is common to all forms of radiant energy and is expressed in the *inverse square law*, which states that the "radiant flux density at any surface varies inversely as the square of the distance of that surface from the source of radiation."

The numerical value of illumination is identical whether expressed in foot candles or in lumens per square foot.

Remember that the equation

$$E = \frac{I}{d^2}$$

can be used to calculate illumination only for a spherical surface with the source of radiation at its center. This condition is known as *normal illumination*.

For small surfaces where the distance from source to surface is large in comparison to the dimensions of the surface, the formula can be used with little error. Over large surfaces where all the flux paths are not perpendicular to the surface, the diffusion of light is not uniform. The flux intensity is reduced at points farther removed from the source by an amount proportional to the cosine of the angle of incidence; therefore, this factor must be included in calculations. The formula

$$E = \frac{I \cos i}{d^2}$$

is much more accurate under these conditions. Even with this formula, however, the surface dimensions must still be small as compared to the distance from the source.

Up to this point, we have considered the luminous intensity of a source in terms of point-sources; therefore, when we refer to luminous intensities of larger surfaces which may or may not be self-luminous, we must use the quantity or term *brightness* to specify luminous intensities of unit areas. Brightness is defined as the luminous intensity of a unit area of a surface in a given direction. Brightness is expressed in terms of candle power per square unit of area. It is generally expressed in terms of square centimeters.

The difference between brightness and illumination can be illustrated by considering this page which you are reading. The page is uniformly illuminated (or nearly so), but the printed letters reflect less of the incident light and therefore are less bright than the white paper upon which they are printed.

Brightness of a surface and the illumination on the surface would be numerically equal

only if the surface reflected all of the light that fell upon it. The table on the right shows the approximate values of brightness of some familiar self-luminous and nonluminous objects.

Generally the brightness of a surface depends upon the direction from which it is viewed, but there are some materials which scatter light in such a manner that their brightness is the same from all angles of view. Examples of such light-diffusing substances are magnesium-oxide and new-fallen snow. For surfaces of this type, a unit of brightness called the *lambert* is used.

A lambert represents the brightness of a perfectly diffusing surface which is emitting or scattering light in the amount of one lumen per square centimeter. When the reflected light is less than the incident light, the brightness in lamberts is equal to the product of the illumination and the reflection coefficient of the surface material.

Sun's disk	153,000,000 cp./sq. ft.
Crater of a carbon arc	14,000,000 cp./sq. ft.
Tungsten lamp filament	465,000 cp./sq. ft.
Moon's disk	465 cp./sq. ft.
Clear blue sky	370 cp./sq. ft.
Newspaper	1.8 cp./sq. ft.

Brightness of common objects

Measurement of Light Intensity

The relative intensities of two or more sources are not discernible to the human eye by direct viewing, but whether or not two surfaces side by side are equally illuminated can be determined accurately. The matching of illumination on two adjacent surfaces is the basic principle of the *photometer*, a device employing two lamps located at some suitable distance apart with a screen located between them. Each side of the screen is illuminated normally by one of the sources. That is, the flux paths from each source are perpendicular (or nearly so) to all points on the surface of the screen.

The screen is moved along the flux path between the two sources until the same degree of illumination is observed on both sides. The distances from the lamps to their respective sides of the screen are then measured. From the equation

$$E = \frac{I}{d^2}$$

you can establish the ratio

$$\frac{I_1}{d_1^2} = \frac{I_2}{d_2^2}$$

where I_1 and I_2 are the luminous intensities of the lamps in candle power and d_1 and d_2 are their respective distances from the screen. Thus, if the value of either I_1 or I_2 is known, the value of the other may be readily computed. The figure to the right illustrates the application of the inverse square law as applied in the foregoing calculation:

If distance d_1 is found to be twice the value of d_2 and if I_2 represents a source of 16-candle power intensity, the intensity of I_1 must be four times that of I_2 to produce the same illumination of the translucent screen.

From the second equation above:

$$\frac{I_1}{d_1^2} = \frac{I_2}{d_2^2}$$

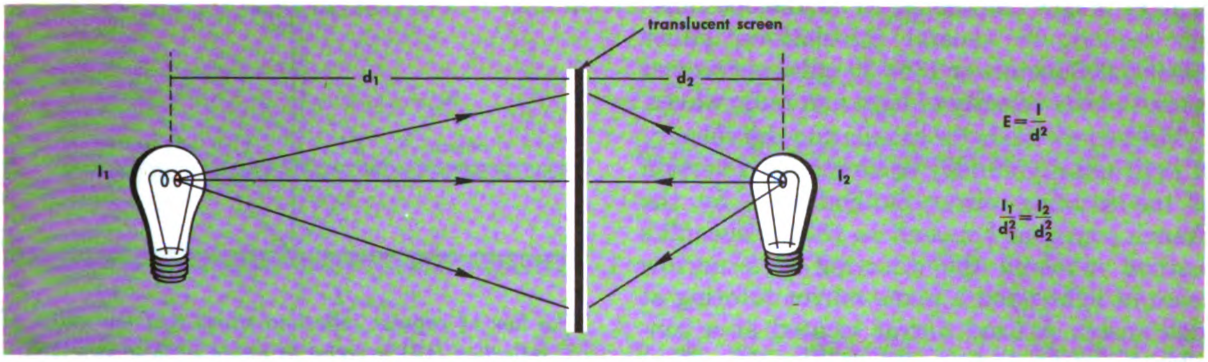
Therefore, if d_1 is equal to twice d_2 and if I_2 is 16 candle power,

$$\frac{I_1}{2^2} \qquad \frac{16}{1^2}$$

Thus, I_1 equals 4 times 16 cp; I_1 equals 64 candle power.

The foregoing method of computing the intensity of one light source in relation to another is satisfactory when both sources are of the same color. A different procedure is necessary if the lamps are not the same color. Lamps of different colors can be matched with a standard lamp by varying the potential on the calibrated lamp. Lowering the potential produces a noticeably yellower color; increasing the potential increases the whiteness of the color.

Another way for comparing lamps of different colors is to use a so-called *flicker photometer*, which employs a rotating prism. The rotating prism enables the viewer to see one side of the screen and then the other alternately in rapid succession. Upon increasing the frequency of alternation, a value will be found for which the flicker due to color



A Photometer is a practical application of the inverse square law

difference disappears. The colors appear to blend into a single hue. If the frequency is not too high, however, the flicker due to illumination difference remains. The photometer screen is then moved until this flicker also disappears. The lamps can then be compared in the usual way.

Photo-electric cells with suitable light filters can also be used to compare the candle power of lamps or other sources of light of different colors. Since photo-electric current is proportional to the illumination on the photo-cathode, the candle power of the lamp under test can be expressed in terms of the current it produces as compared with that produced by a calibrated standard lamp. The following figure illustrates how a photo-emissive cell may be utilized in this application.

Light from the source (lamp, etc) falls upon the cathode of the cell, which is coated with some light-sensitive material. This material releases or emits electrons when struck by the incident light rays. These electrons are attracted to an anode which has been placed at a positive potential by means of a suitable battery. A meter connected in series with the cathode, anode, and battery registers the increase in current produced by the incident light. Filters may be placed between the source and the cathode so as to permit only light of a desired wavelength (color) to reach the cathode.

In selecting reference stars for celestial-navigation systems, we are concerned primarily with evaluating intensity in one direction only. Photometric devices as mentioned in the preceding paragraphs could be adapted to this application.

VELOCITY OF LIGHT

Early experiments to determine the velocity of light gave inaccurate results because of time losses in the operation of the equipment used. But recent observations using improved optical devices and electronic timing controls have greatly reduced the error. The velocity

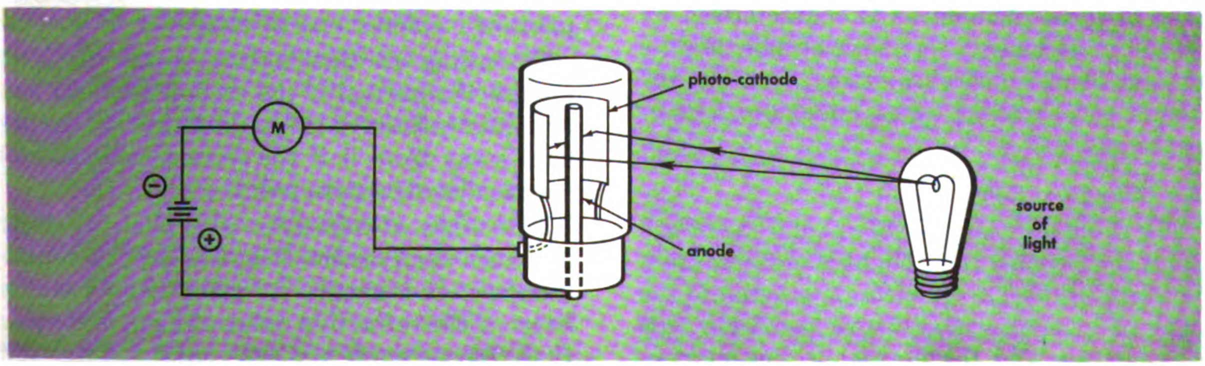


Photo-emmissive cell used to measure light intensity

of light is now considered to be 186,300 miles per second. In general calculations, either 300,000,000 meters or 186,000 miles per second is used as the velocity of light and radio waves.

Modern methods for calculating the velocity of light usually employ some modified version of the rotating-mirror method used by the French physicist Foucault in 1850. Foucault's method involved the directing of a narrow beam of light upon a plane mirror rotating at high speed. A fixed mirror located at a considerable distance from the flashing (rotating) mirror received the momentary flash of light and reflected it back to the rotating mirror from which it was again reflected. During the interval of time required for the light beam to travel through the measurable distance between the rotating mirror and the stationary mirror and back, the flashing mirror rotated through a definite angle.

From the angular velocity of the rotating mirror and the distance between the mirrors, the time required for the light to travel from the flashing mirror to the fixed reflector and return was computed, and from these factors the velocity of the light was determined.

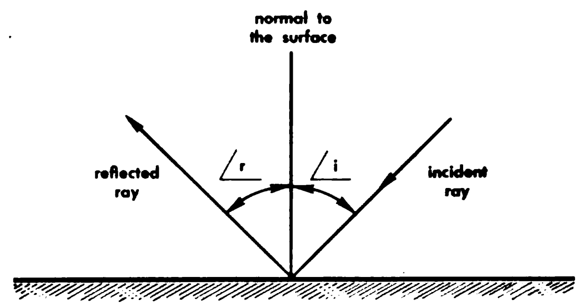
Foucault's method was improved by an American physicist, Michelson, who utilized a rotating octagonal mirror, from one face of which light from a source of high intensity was reflected to a distant plane mirror. The returning light ray from the fixed mirror would strike another face of the rotating mirror from which it was reflected into the observer's telescope.

REFLECTION AND REFRACTION OF LIGHT

Two of the most important phenomena of geometric optics are *reflection* and *refraction*. Each of these phenomena is characterized by a basic relationship or natural law and is present in all optical systems such as the human eye, lenses, prisms, telescopes, microscopes, etc.

Law of Regular Reflection

The fundamental law of regular reflection states that when a ray of light is reflected from a surface, the angle of reflection is equal to the angle of incidence; and the reflected ray, the incident ray, and the normal (a line perpendicular to the reflecting surface at the point



The law of regular reflection states that the angle of reflection (LR) is equal to angle of incidence (CI)

of reflection) all lie in one plane. The accompanying illustration demonstrates this law.

This law of reflection was found to apply when light was reflected at the interface (common boundary) between two unlike media such as air and a solid surface. The one medium in this case was transparent (air) and the other (solid) was opaque.

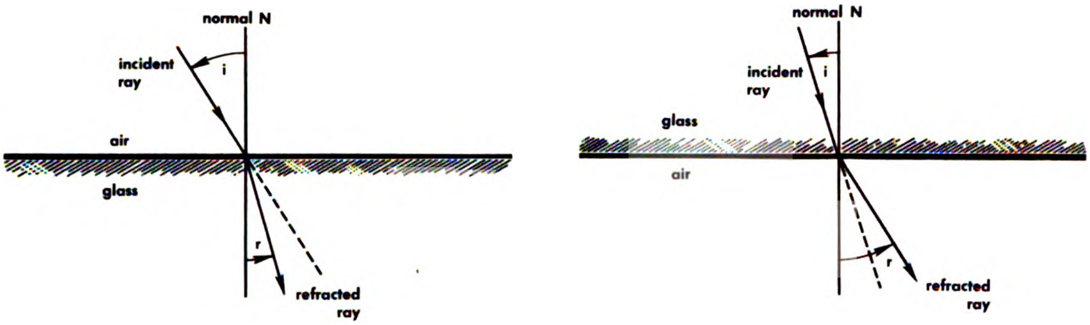
Refraction of Rays

If the second medium is not opaque, some of the light passes on through it and is refracted or bent in a direction which generally is different from the direction of the incident rays, but the refracted rays still lie in the same plane as the normal to the surface. The degree to which this bending action takes place determines the *refractive index* of the medium.

Each refracting medium has a specific refractive index in any one refracting medium or, in different words, there is a constant ratio between the sine of the angle of incidence and the sine of the angle of refraction, in any one refracting medium. Both angles are measured with respect to the normal, a line, in this case, perpendicular to the reflecting surface at the point of incidence and refraction.

According to Snell's law, a ray is bent toward the normal if the second medium has a greater refractive index than the first and is bent away from it if the second medium has the smaller refractive index.

The angle of refraction is smaller for some mediums than others. For example, the angle of refraction is smaller for glass than for water because the refractive index of glass is greater than that of water and thus tends to bend the refracted ray nearer to the normal.



Refraction of light rays

The figures above illustrate examples of refraction. They show how a light ray is bent in passing from one medium to another. The figure on the left illustrates the refraction of a light ray in passing from air to glass, and the right-hand figure illustrates the refraction of the ray in passing from glass to air.

You can conclude from the above information that a ray will deviate toward the normal when its velocity is decreased, and it will deviate away from the normal when its velocity is increased. When a ray passes from one medium into another which tends to reduce its velocity to a greater degree than it was reduced in the first medium, it deviates toward the normal. It deviates away from the normal if the ray passes from a medium in which it met more opposition (velocity slowed) into another medium in which it encounters less opposition.

As noted above, the ratio of light velocities in two mediums which are in contact is a constant for those two mediums. We also noted above that this ratio is referred to as the *refractive index* of the second medium with respect to the first. The refractive index for two mediums is represented by the symbol $\mu_{1,2}$ (mu sub 1, sub 2), with the order of the subscripts indicating the direction of the ray of light.

The *law of refraction* states that "when a wave travels obliquely from one medium into another, the ratio of the sine of the angle of incidence to the sine of the angle of refraction is the same as the ratio of the respective wave velocities in these mediums and is a constant for two specific mediums."

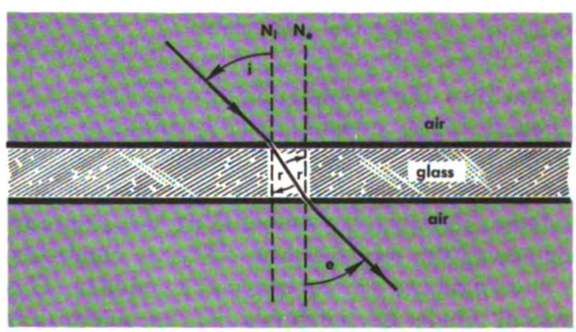
$$\frac{\sin i}{\sin r}=\frac{V_1}{V_2}=\mu_{1,\ 2}$$

The absolute *refractive index* of a medium is its index compared to the refraction of light in a vacuum, which is considered as unity (1.000). The refractive index of air is so small that for practical purposes it is used as a standard.

Another important phenomenon of refraction which you will encounter in the study of celestial-navigation systems is that when a ray of light passes through one or more parallel-sided media and emerges into the original medium, it is displaced laterally but its direction is unchanged.

In the following figure, "i" represents the angle of incidence, "r" the angle of refraction, and "e" the angle of emergence of the ray of light passing through the parallel-sided medium, glass.

A ray of light passing from air through a parallel-sided pane of glass and emerging into air follows the law of refraction for each surface of the glass. Thus, angles "i" and "e"



Passage of light from air to glass to air

are equal, and the incident ray is parallel to the emergent ray. Also,

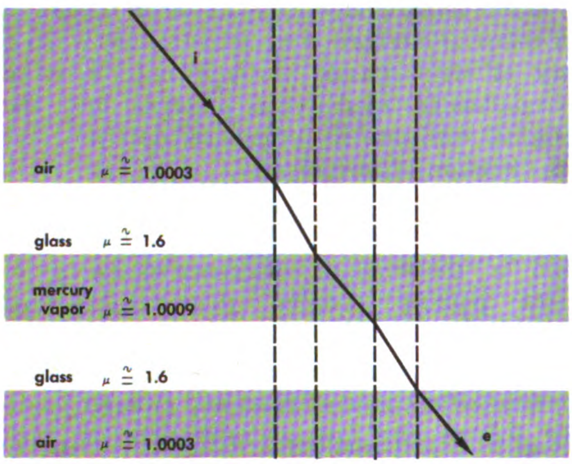
$$M_{ag} = \frac{1}{\mu_{ga}}$$

which shows that the refractive index of glass (g) with respect to air (a) is the reciprocal of the refractive index of air with respect to glass,

These same principles hold true for a ray of light passing through more than one parallel-sided medium, as illustrated in the following figure. The figure shows how light is refracted in passing through several parallel-sided media. The refractive index of each medium is noted in the figure.

Notice in the following illustration that the angle of refraction in the glass is less than in the air or gas.

A ray of light, passing through a medium with a high refractive index toward one with a lower refractive index, is refracted if the angle of incidence is not too large when it



Refraction of light in passing through several parallel-sided media

passes into the second medium. If the ray is inclined to an ever-increasing angle of incidence, it arrives at some position at which it no longer passes into the second medium. Instead, it is totally reflected at the common surface of the two mediums. Such a condition is known as the *critical angle* of incidence. The critical angle of incidence is the maximum angle at which the ray of light striking the surface of a medium passes through it.

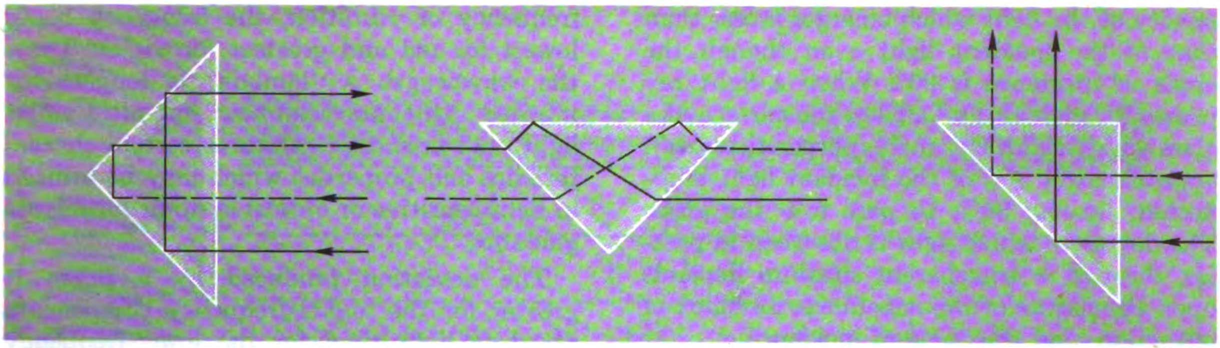
When this critical angle of incidence is exceeded, the refracted ray grazed the surface with an angle of refraction of 90° in the medium, and the ray is totally reflected from the surface of the medium. This principle is often applied in optical instruments. Prisms are used to achieve "total reflection" whenever it is desirable to avoid the use of silvered mirrors.

Basically, a *prism* is a transparent body bound in part by two plane faces which are not parallel. The line in which these faces meet is called the edge of the prism. Prisms are taken up in more detail later.

The figure on the right shows three positions of a prism, which has 45° angles, in the path of light rays. The first position shows the prism functioning like a plane mirror tilted downward; the second position shows how the prism inverts an image; the third position shows how the prism displaces the rays and reverses their direction.

It is possible to compensate for the displacement of a ray of light passing through one substance. The displacement is compensated for by causing the ray to emerge from the displacing substance into another material which has a refractive index that is the reciprocal of that of the first substance. This procedure tends to cause the final angle of emergence to equal the original angle of incidence. Such compensation may be necessary in cases where the desired light must pass through some secondary medium, such as a pressurized chamber, before it reaches the receiving device. To present the light to the receiving device in its original angle, the displacement produced by the secondary medium must be nullified. Problems similar to this example exist in the case of optical instruments used for celestial observations. The instruments must be housed within some protective container, such as the hull of a missile. The original light ray passes through a heavy glass pane and then perhaps into a zone of a different refractive index than the external atmosphere.

Complex computations may be involved since the refractive indexes of various media vary with changes of temperature and density. For example, the μ of water is approximately



Total-reflecting prism in varying positions

1.333, but if frozen into ice, its μ is approximately 1.31.

Variations in atmospheric conditions produce variations in the refraction of light from celestial bodies; therefore, in selecting a star as a *fix* for a celestial navigation system, it is desirable to choose one whose light has an angle of incidence as nearly equal as possible to the normal for the earth's atmospheric layer.

Light rays passing through successive layers of air of varying densities or through translucent media with varying indexes of refraction are bent or directed along a new path of radiation by each layer. When this process occurs at a visible rate, it produces an effect called *scintillation*, more commonly called glitter or twinkle. The twinkling of starlight is produced in this manner. A similar effect can be detected with respect to electromagnetic waves which pass through zones of conductance which vary in refractive characteristics for each specific wavelength of the radiations.

Light reflected from surfaces with many facets, such as cut diamonds, tends to produce the effect of flashing or scintillating. Electromagnetic waves, reflected from various objects and arriving at the receiver at intervals slightly later than the direct beam, tend to produce phase differences and frequency in varying degrees. This results in phase cancellation and interference, referred to as "fading," "swinging signal," or "ghosts."

In television, a "ghost" image on the picture screen occurs when the transmitted signal arrives at the receiver along more than one path, the paths being unequal in distance. This condition produces two or more images slightly displaced in phase and time.

Radar signals are subject to the same effect. The effect is present when the signals are reflected from moving surfaces or from more than one object, or when they are refracted by inequalities in terrain or atmosphere. To minimize this effect of fluctuation or scintillation, the radar beam is narrowed so that both the incident and reflected beams travel through a narrow zone and are less subjected to reflection from objects other than those directly along the principal axis of the beam. When scintillation is produced by a fixed or constant succession of media, the frequency of the scintillation may be used to identify the path or locate the source of the radiation.

Functions of Prisms

A simple optical device which may be used to compensate for light displacement or to produce some specific deviation in a beam of light is the prism, which was described briefly in the preceding section.

The amount of deviation produced by a triangular prism depends upon the angle of the prism, the refractive index of the material of the prism, and the angle of incidence of the light passing through the prism.

By controlling the position of a common triangular prism with respect to the source of light and by choosing a prism of proper angle, the desired deviation of the light beam can be produced even to the extent of producing total reflection.

Some light is absorbed or reflected by the transparent medium; consequently, the in-

tensity of the light which emerges is less than that of the incident light even though the angle of incidence is 90°. When a light ray passes through air and strikes the surface of a glass window at a 90° angle (perpendicularly), approximately 4% of the incident light is reflected, leaving 96% to pass into the glass. The degree to which light is absorbed in passing through a transparent medium depends upon the nature of the substance, its index of refraction, and the angle of incidence.

Spectra

The "rainbow" of color produced by sunlight passing through a crack into a darkened room is a visible color spectrum. This phenomenon is the result of refraction and dispersion of light rays, and it exemplifies the basic principle of spectrography.

Actually, white light is composed of light rays of many hues blended together. When such light is passed through a narrow opening or slit into some diffracting medium, rays of each different color are diffracted at different angles and spread out into merging bands of six principal colors. These colors are red, orange, yellow, green, blue, and violet. The diffracting of radiations of light at different angles produces a "rainbow" or spectrum of light. The prism shown below is acting as a as a diffracting medium, thus producing a spectrum of light.

A spectrum is composed of hundreds of hues which are grouped broadly into the six principal colors mentioned above. These colors appear in the order of increasing deviation from red to violet.

The color of light is determined by its frequency of vibration, the frequency being lowest for red and greatest for violet. Thus, the wavelength λ (lambda) is longest for infrared light and shortest for ultra-violet light, and *each color* possesses certain characteristics and properties which may be utilized in light-sensitive or optical systems for missile guidance.

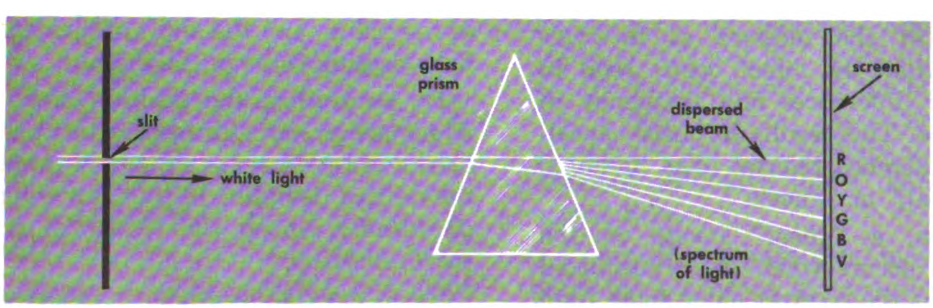
The deviation produced by a prism is greater for the higher-frequency components of white light than for the lower-frequency hues, but no definite relationship exists between frequency and deviation. Thus, prisms of different materials spread the component colors of the spectrum in varying degrees.

Some materials produce *anomalous* dispersion; that is, some prisms are composed of materials that do not disperse white light in the regular sequence of colors. Such prisms deviate certain colors to a much greater degree than others and also absorb certain portions of the spectrum.

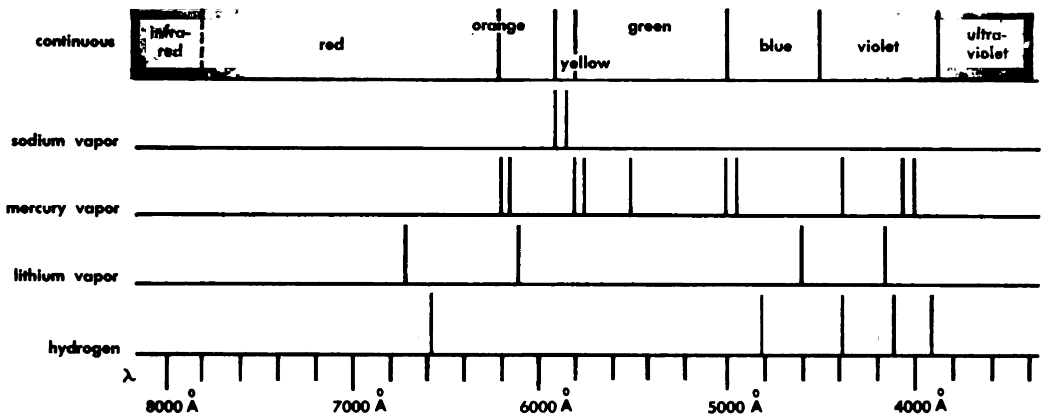
In a quantitative study of the spectrum, it becomes necessary to refer to each particular part of it with definiteness. This is done by specifying any hue by the vibration rate of the light source and its corresponding wavelength. As in all other forms of wave motion, the velocity of light is equal to the wavelength times the frequency,

$$V = f\lambda,$$

where "f" equals the frequency and "λ" the wavelength. When light is retarded by a medium such as glass, the frequency is unaltered; but, since velocity decreases, λ must also decrease in direct proportion. Therefore, λ is not a constant quantity for a given vibration, but depends upon the medium.



Dispersion of light through a prism



Continuous and bright-line spectra

In general, values of λ for different colors of a spectrum are given as the wavelength in air or in a vacuum. Wavelengths are so short for visible light that a special unit of length shorter than the centimeter is commonly employed to measure them. This unit is known as the *Angstrom unit*, named in honor of a Swedish physicist. The Angstrom unit (A) is equal to one one-hundred millionth of a centimeter (10^{-8}cm).

In some cases of lower-frequency radiations, the micron may be used as a measuring unit. One micron is equivalent to 10,000 *A*.

Spectra are often classified into three general types: emission, absorption, and solar spectra.

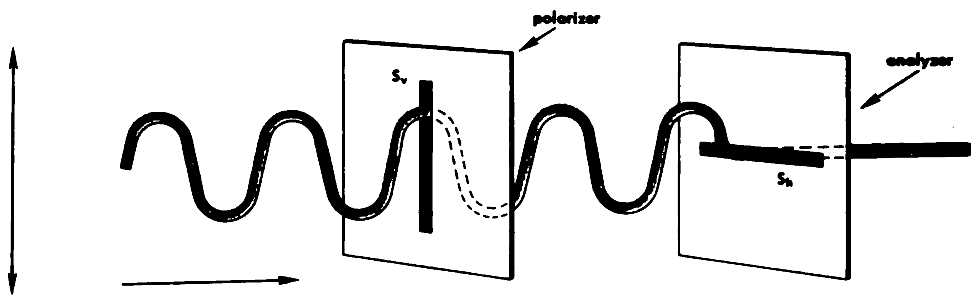
EMISSION SPECTRUM. A spectrum produced by a glowing object is termed an *emission spectrum*. Its appearance depends primarily upon the composition and state of the luminous object.

Incandescent solids and liquids produce continuous spectra, extending from color to color without interruption. Luminous gases and vapors yield spectra consisting of definitely placed bright lines. Each bright line is an image of the slit through which the radiation is received.

Every gas emits radiation of particular wavelengths, and each spectrum is characteristic of the radiating substance. For example, sodium vapor yields two bright lines in the yellow part of the spectrum, while mercury vapor yields several bright lines, the most conspicuous being in the green and blue regions. The illustration above shows the bright-line spectra of several elements over the range of the visible spectrum. The continuous spectrum at the top is an uninterrupted series of images of the illuminated slit.

The number of lines in a bright-line spectrum depends upon the amount of energy with which the atoms of the source are excited to produce glowing, as well as being dependent upon the nature of the source. The greater the excitation of the atoms of a substance, the greater the number of lines that appear in the spectrum of the substance.

ABSORPTION SPECTRUM. *Absorption spectra* occur when white light has passed through an absorbing medium before the light is dispersed. A glowing solid, or other source from which the light radiates, yields a continuous spectrum; but when the light passes through the absorbing medium, radiations of particular wavelengths are absorbed. The resulting spectrum is usually crossed by dark spaces because of the absence of the absorbed radiations. If the absorbing material is solid or liquid, these dark spaces appear as broad, structureless bands. If the material is gaseous, the dark spaces consist of dark lines which occupy the same positions as the bright lines in the corresponding bright-line spectrum.



Analogy of polarization of a transverse wave

SOLAR SPECTRUM. A spectrum formed by radiations from the sun is called the *solar spectrum*. This spectrum appears continuous from a casual inspection; however, a more critical examination shows that it is crossed by numerous dark lines. No doubt the sun's radiation comprises all wavelengths in the visible range, but in passing through the sun's "atmosphere," certain wavelengths are absorbed. Thus, the spectrum observed is in reality an absorption spectrum of the sun's atmosphere.

We now can briefly outline the types of spectra. Continuous spectra are produced by light from incandescent solids and liquids. Bright-line spectra come from incandescent vapors or gases. Absorption spectra are produced by light passing from an incandescent solid or liquid through an incandescent vapor or gas. And we found that the solar spectrum is actually an absorption spectrum.

POLARIZATION OF LIGHT

When the question arises as to whether light waves are longitudinal like sound waves or transverse like elastic waves, it is advisable to consider the phenomenon of *polarization* of light.

If a beam of light is passed through a substance composed of two colors, such as a crystal of tourmaline or a sheet of polaroid, the beam's passage is restricted to a particular plane of vibration. The beam is said to be *plane-polarized* and will pass through a second crystal of tourmaline or sheet of polaroid only if the tourmaline or sheet is oriented exactly the same as the first. If the second crystal or sheet is rotated 90° with respect to the first, no light passes through.

The first crystal or sheet is called the *polarizer*, and the second the *analyzer*. The analyzer is so named because the angle of polarization may be determined from the angle through which the analyzer must be rotated between the points of light passage and cutoff. The above figure illustrates the principle of polarization of a transverse wave. The vibrating rope corresponds to a beam of light in this analogy.

Vertical vibrations of the rope pass through a vertical slot (S_v), but they are stopped by the horizontal slot (S_h). The wave represented by the rope is polarized; that is, it is vibrating in one plane only (in this example the vibrations are in the vertical plane). Light can be similarly polarized by certain optical substances such as tourmaline crystals and "Polaroid," which is the trade-name for a commercial product possessing this property.

If slot S_h were rotated 90° in the foregoing analogy, the waves would pass through it. S_v limits the wave motion to the vertical plane and is the polarizer. S_h is the analyzer. Since S_h must be rotated 90° to permit the waves to pass, it is evident that the degree of polarization produced by S_v is 90°. The degree could be determined by the required distance of rotation of S_h between points of passage and cutoff, even though S_v were not visible.

Both theory and experiment have shown that longitudinal waves cannot be polarized; therefore, we may conclude that light must have a transverse wave motion if it has the

form of wave motion at all.

Light can be polarized to a considerable extent by reflection alone. All reflected light is polarized to a certain degree because of scatterings caused by dust and vapor particles. Scattering is more apparent in light of short wavelengths. This statement is supported by the color changes of the sky. The mid-day sky appears blue, but a sunset appears reddish. The sunset appears reddish because the light travels through a longer path of the earth's atmosphere, resulting in the short-wavelength blue light being dissipated by scattering. This scattering leaves a predominance of the longer-wavelength reddish hues.

FLUORESCENCE AND PHOSPHORESCENCE

Some natural substances possess the property of emitting light when excited by an external force, such as "bombardment" by electrons or certain forms of radiant energy. In some cases, light is emitted by the substance only while the bombardment is taking place; in other cases the emission may persist for some interval of time after the external excitation has ceased. These properties are termed *fluorescence* and *phosphorescence* respectively, and the duration of light emission is referred to as the *persistence* of the material. Phosphorescence is commonly referred to as *afterglow*.

One of the fluorescent substances commonly used in electronic application is willemite (zinc orthosilicate Zn_2SiO_4), a crystalline zinc compound varying in color from white, greenish-yellow, and green to shades of red and brown. The white and green varieties are frequently used in the coatings on the screens of cathode-ray tubes used in oscilloscopes and radar indicators. Willemite is used in television kinescopes in combination with other substances which impart the desired degree of persistence.

Generally when the emission is induced by some form of radiation, the luminescent substance emits light which is of a longer wavelength than the incident radiation. This phenomenon is encountered in the fluorescent lamp in which certain phosphorus compounds are excited by ultraviolet radiation which is invisible. The compounds then emit visible light of various colors. The fluorescent compounds used in the "soft white" and "daylight" lamps are combinations of zinc beryllium silicate and magnesium tungstate.

Infrared radiations reflected from objects onto a photo-sensitive surface which is sensitive to infrared light may be utilized for seeing in the dark. A device called the "snooperscope" was developed during World War II for this purpose.

Essentially, the snooperscope consists of a source of infrared light and an image tube having a light-sensitive cathode. An infrared lamp projects a beam from which all visible light is filtered out. The reflected rays from the beam are caught upon a cesium cell, which is highly sensitive to infrared light and which forms the cathode of the image tube. The reflected radiations incident upon the cell cause it to emit photo-electrons which are then focused on a fluorescent screen to form a visible image, as in the cathode-ray oscilloscope.

You will find in later chapters that a knowledge of the emission spectrum of any radiating surface is of value in designing a homing system to guide a missile toward a target presenting such characteristics of emission. Such a knowledge also is of value in designing a navigation system employing light from certain fixed stars as a reference.

Obviously, such systems must be highly sensitive and selective and thus require the use of optical components of the highest precision. Marked progress has been made in improving basic optical instruments and in adapting them to new applications, and much progress has been made in developing electronic systems which employ the basic principles of optics. This progress has created systems which are extremely sensitive and selective with respect to heat, light, and other forms of radiant energy.

Many of the present-day missile guidance and navigation systems are of these improved types, and further research and development undoubtedly will produce revolutionary improvements in their accuracy and reliability as well as in their adaptation to industrial and commercial usage.

How Optical and Electronic Principles Work Together

A brief discussion of some of the basic optical and electronic components is given here to help you in your understanding of the guidance and navigation systems of missiles. In considering these basic components, you will gain a general idea of how light waves are utilized along with electromagnetic waves in electronic equipment. You will learn about some of the properties common to both types of waves. Devices used to utilize radiant energy (light) are taken up first.

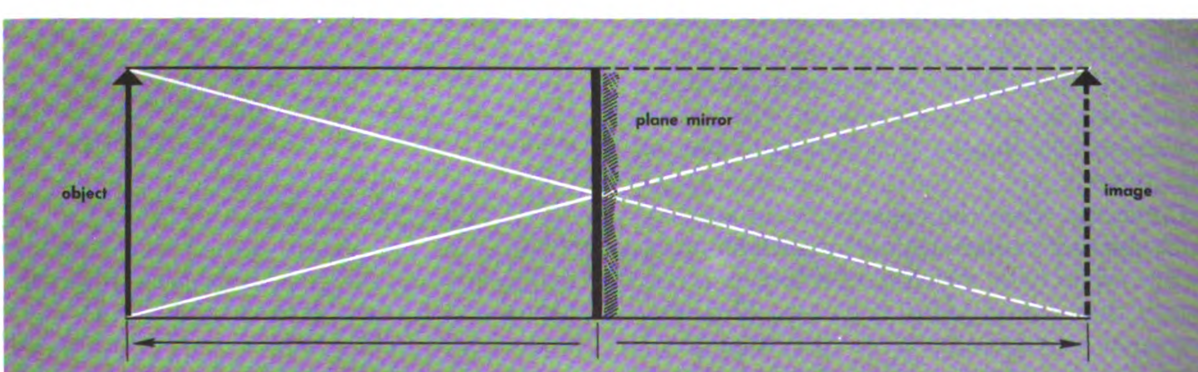
SIMPLE OPTICAL DEVICES

To utilize the various phenomena of radiant energy, simple optical devices such as mirrors, prisms, and lenses are employed either singly or in combinations to suit the complexity of the requirements.

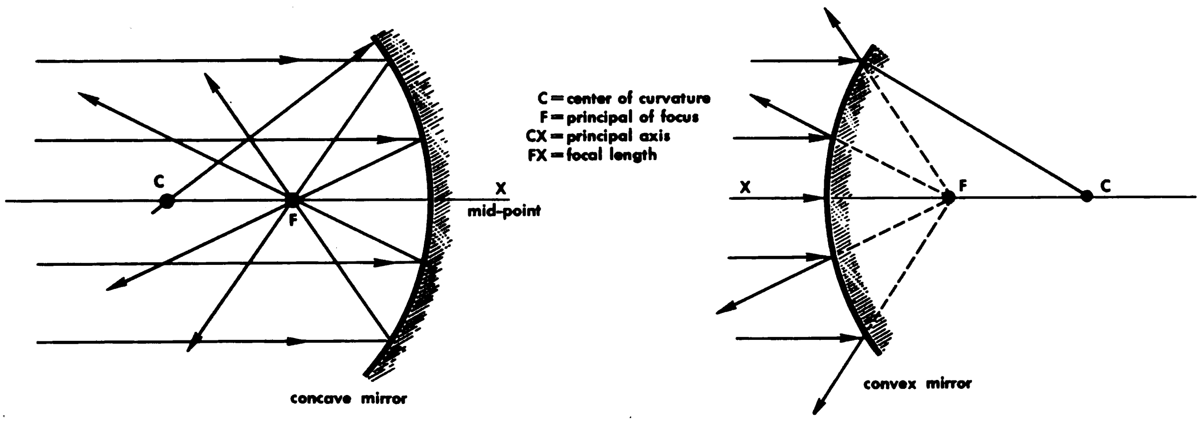
Functions of Mirrors

A *mirror* is a polished or smooth surface that forms images by means of reflected light. When an object is placed before a plane mirror, a right-side-up image is formed. The image appears to be just as far behind the surface of the mirror as the object is in front of it. You encounter this simple phenomenon whenever you approach a mirror. You note that your image appears to move toward you from a point to the rear of and equally distant from the mirror surface. This phenomenon conforms to the laws of reflection and is an important one in regard to optics.

The figure below shows a plane mirror with an image that is direct and vertical and appears to be as far behind the mirror as the object is in front of it.



Object distance from a mirror equals image distance



Reflection of incident light by spherical mirror

Curved reflecting surfaces may be designed to give variations in the apparent position and dimensions of the image by deviating a beam of light, causing it to be more or less converging when reflected than it was when incident upon the mirror.

Spherical mirrors are classified as concave or convex. A concave mirror has its reflecting surface on the inside of the spherical shell while the convex mirror has its reflecting surface on the outer side of the shell.

The center of the spherical surface is called the center of curvature of the mirror. A line from the middle point of the mirror surface to the center of curvature is called the principal axis of the mirror. The above figures illustrate the two types of spherical mirrors and how they reflect incident light.

Point "F" is the principal focus of the mirror. The distance of the principal focus from the mirror is the focal length. The principal focus of a spherical mirror is located on the principal axis half-way between the center of curvature and the mirror surface.

A concave mirror is a converging reflector because it actually converges the light rays. A convex mirror is a diverging reflector because it makes the light rays appear to diverge.

The ratio of the dimensions of a mirror with respect to the radius of curvature is referred to as the aperture of the mirror. Most optical mirrors are of small aperture and produce only a slight inclination of the incident rays with respect to the principal axis.

The ratio of the size of the image to the size of the object is referred to as the magnification of a mirror. A spherical convex mirror always causes the image to appear reduced in size but right side up. A spherical concave mirror can be made to produce *inverted* images which appear to stand out in space or to produce right-side-up images which appear to be behind the mirror.

The inverted images are called *real images* because they appear to exist where the rays of light are focused and because they can be localized upon a screen. Right-side-up images are called *virtual images* and cannot be projected upon a screen.

Mirrors are part of the optical systems used in navigation and in surveying instruments. Often an optical system is used in conjunction with an electronic system which converts the optical data into electrical data. The electrical data can then be transmitted over great distances. Television is an example of such a combined optical and electronic system.

Uses of Prisms

Prisms — discussed previously in the section on light dispersion, reflection, and refraction — are used in binoculars, spectroscopes, refractometers, and many other optical devices which utilize the phenomena of light dispersion and refraction.

The characteristics of light and the chemical composition of light sources may be determined by methods which involve the use of prisms.

In celestial navigation systems, the light from the reference star can be passed through a prism in such a manner that either the predominant color or any desired color present in

the dispersed beam can be utilized to activate a suitable photo-cathode or light-sensitive cell. The photo-cathode or cell produces and maintains a voltage or current-output level which is proportional to the frequency (color) and/or intensity of the light which falls upon it.

Possibly now you can visualize how such a constant voltage or current can be used to control a navigation system along a fixed path which is referenced to one or more specific light sources, such as stars. Many of the fixed stars emit light which is characterized by some specific color, such as Arcturus, which is blue, and Aldebaran, which is red. Spica is a spectroscopic binary; that is, Spica's white light is a blend of two predominant colors.

It is possible to identify a star by its light spectrum once the spectrum has been tabulated. Spectroscopic equipment involves the use of prisms for the dispersion of the reflected light from the planet or the emitted light from the star.

Composition and Uses of Lenses

Usually, in addition to mirrors and/or prisms, lenses in some form are found in both optical and electronic-optical systems.

An optical lens is basically a piece of glass or other transparent material which has two opposite regular surfaces. Both surfaces may be curved, or one surface may be curved and the other plane.

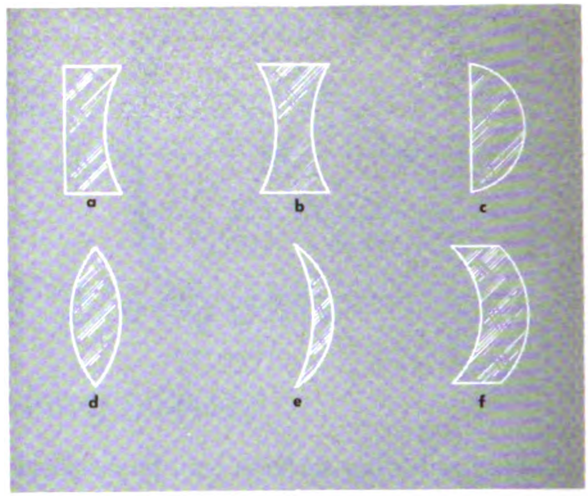
Lenses are used singly or in combination with other lenses, prisms, or mirrors to perform specific functions. The primary function of a lens is to form an image by changing the direction of the rays of light. Such optical instruments as eye-glasses, cameras, microscopes, and telescopes are basically systems of lenses.

You will find that the curved surfaces of lenses generally are spherical, although in some rare instances cylindrical lenses may be encountered.

Spherical lenses may be classified broadly into six types:

a. Plano-concave: one plane surface and one opposite concavely curved surface.

b. Double-concave (biconcave): two opposite concavely curved surfaces.



Six types of spherical lenses

c. Plano-convex: one plane surface and one opposite convexly curved surface.

d. Double-convex (biconvex): tow opposite convexly curved surfaces.

e. Converging concavo-convex (converging meniscus).

f. Diverging concavo-convex (diverging meniscus).

In all of these spherical lenses, the line joining the centers of curvature of the two surfaces is a line of symmetry of the lens and is called the *axis of the lens*.

A lens whose focus for parallel rays is *real* is called a converging lens, as in the case of mirrors. And a lens which has a virtual focus for such rays is called a diverging lens.

Many of the principles of lenses which apply to light also hold true for other forms of radiation such as electromagnetic waves, cathode rays, etc.

The first outstanding scientific application of the principles of optics was in the field of astronomy. A telescope developed by the Italian physicist Galileo Galilei at the beginning of the 17th century is generally regarded as the forerunner of optical instruments; however, it is probable that eyeglasses were used prior to Galileo's time and most certainly some of the properties of simple lenses and prisms had been known for centuries.

SIMPLE MICROSCOPE. A simple magnifier or microscope usually consists of a single bi-

convex lens which converges the light rays in such a manner that a larger image is produced on the retina of the eye than is produced by the object without the aid of the lens. The figure below represents a bi-convex lens serving as a simple microscope or magnifier. The figure shows how the lens produces a virtual image when the object is placed between the focal point and the lens. The image appears on the same side of the lens as the object.

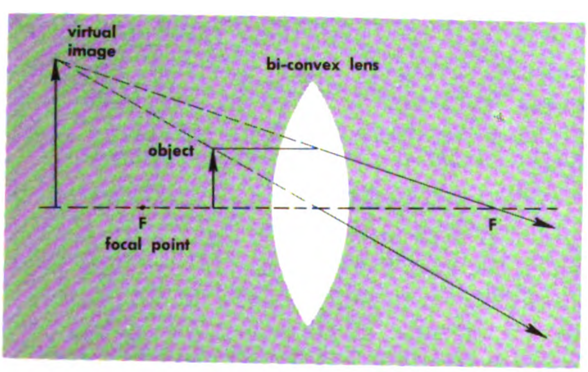
When the lens is positioned at a distance from the object less than the focal length of the lens, the resulting image is *virtual* (right-side up) and on the same side of the lens as the object.

If the lens is positioned far from the object, the image is *real* (inverted) and appears in front of the lens on the opposite side from the object.

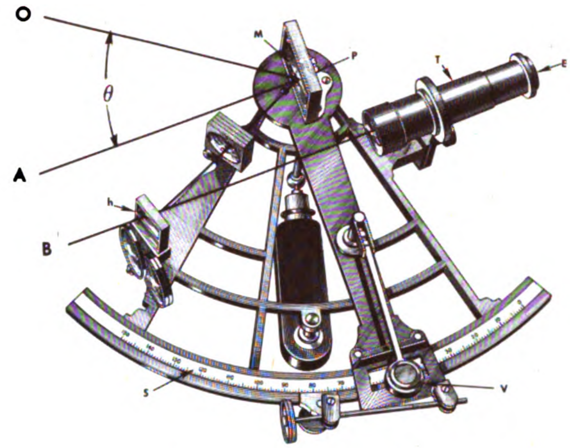
In optical systems employing more than one lens, the image formed by one lens may be the object for another lens. Thus, high magnification of the object may be attained. This is the fundamental principle of complex microscopes and telescopes.

Sextants. Sextants, used in navigation to measure the angle between the sun and the horizon from the position of the observer, make it possible for observers to determine the latitude of their positions.

The sextant illustrated to the right above figure is a telescopic instrument which is held in the hand of the observer while he sights through the telescope. While sighting, he adjusts it until the images of the sun and the horizon are coincident in the field of view.



Simple miscroscope or magnifier



Navigator's sextant

The sextant utilizes two mirrors, one of which is called the *index* glass (M) and the other the *horizon* glass (h). These mirrors are supported perpendicular to the plane of the sextant on a frame to which the telescope and scale are attached. The horizon glass is fixed in position and has one-half of its surface unsilvered so that the observer, looking through the eyepiece (E) of the telescope (T), can see the horizon without reflection. The index mirror (M) is mounted on a movable arm pivoted at P and has a graduated vernier scale (V) to permit the position of the index glass to be read accurately on scale "S."

If the movable arm is positioned so that mirror "M" is parallel to mirror "h," the observer sees the horizon by way of path "AMhE." This image blends with the image observed along the direct path "BE." For this position of mirror "M," the vernier "V" reads zero.

When "shooting" (viewing) the sun at an angle θ degrees above the horizon, the arm upon which mirror "M" is mounted has to be turned through an angle of $\theta/2$ degrees so that the rays from the sun along path "OM" will be reflected along path "Mh" and then reflected into the telescope by the silvered half of the horizon glass (h). The image of the horizon and the image of the sun then coincide, permitting the altitude of the sun to be read on scale "S."

Since the vernier arm is turned through only one-half the angle between horizon and sun, scale "S" is calibrated so that the angular

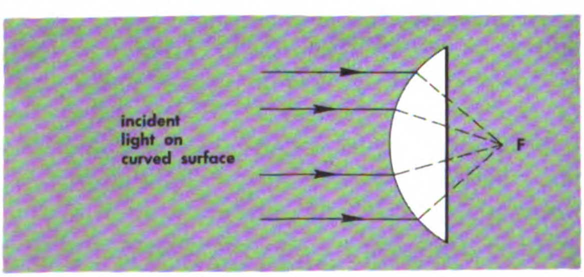
reading is doubled; that is, one-half degree on scale "S" is marked as one degree. Thus the full scale, calibrated from 0° to 140° as shown, actually occupies an arc of 70° magnitude. This arrangement permits direct reading of the angle θ on the scale without having to multiply the indicated angle by two.

Such instruments as astronomical telescopes, engineer's levels, surveyor's transits, and theodolites are essentially telescopes mounted on suitable frames so that the telescopes may be rotated and elevated as required. The angles of elevation and rotation are read on accurately calibrated scales which are referenced both horizontally and vertically by "bubble levels," solar angle indicators, "plumb bobs," magnetic compasses, or other devices employed individually or in combination.

The accuracy of such instruments depends upon the quality of the lenses and other optical components and the precision of the mechanical construction and calibration.

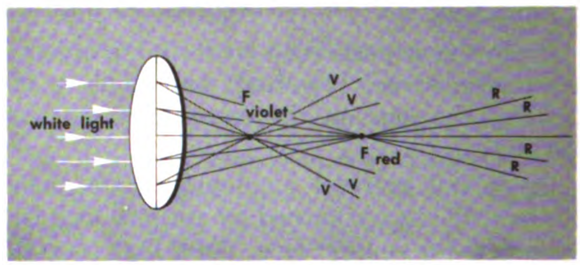
ABERRATION OF RAYS PASSING THROUGH LENSES. The glass from which lenses are made must be free from impurities and must be homogeneous in chemical structure. And the radii of the spherical surfaces must be properly selected. These factors are important in minimizing spherical aberration (deviation of rays from a focal point) to the greatest possible extent. Spherical aberration occurs when rays of light parallel to the principal axis of a lens do not all converge at a common focal point but intersect, instead, along the principal axis at various points. The intersecting of rays produces a blurred image.

A plano-convex lens used so that the incident light falls upon its curved surface produces little spherical aberration, as illustrated in the accompanying figure.



Plano-convex lens with little spherical aberration

Chromatic aberration is another problem encountered in the use of lenses. Chromatic aberration occurs when the various colors present in the incident light converge at individual points along the principal axis of the lens. This is due to the variations in the refractive indexes of the colors. As shown previously in reference to the light spectrum, violet color has a greater refractive index than red; therefore, the focal length of the lens is less for violet light than for red. Thus, the violet rays converge at a point closer to the surface of the lens than the point at which the red rays converge. This phenomenon is illustrated in the figure below.

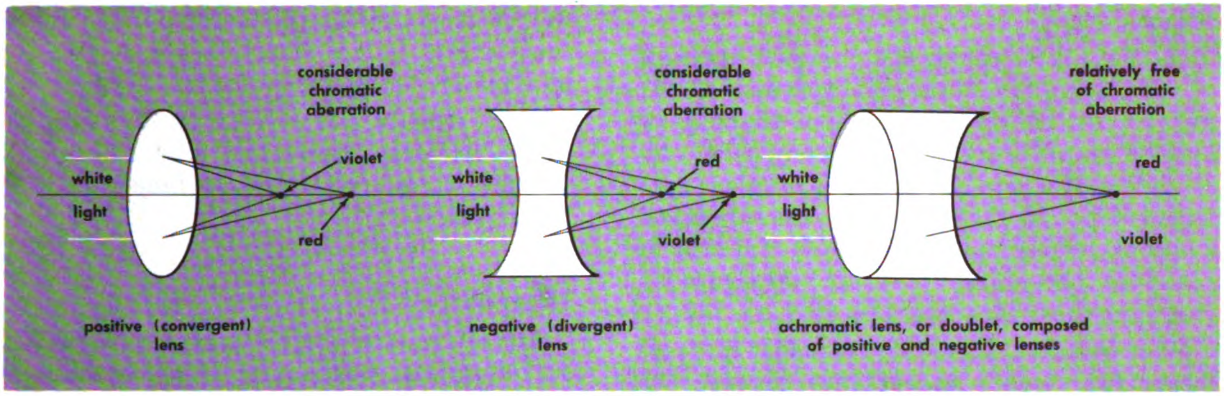


Bi-convex lens showing chromatic aberration

Thus, in passing through a lens, white light tends to disperse into its component colors which in turn tend to focus at different points along the principal axis of the lens. This phenomenon produces chromatic (color) aberration.

To overcome this effect, two or more lenses may be combined so that the divergence produced by one lens is nullified by the convergence of the other. Such a compounded lens is called an *achromatic lens* or *doublet* and generally consists of two lenses with opposite dispersion characteristics placed in contact with each other. The figures on the next page illustrate a bi-convex lens (positive), a biconcave divergent lens (negative), and two lenses combined to form an *achromatic lens* which is relatively free from chromatic aberration.

Indistinctness of an image produced by a spherical lens may result from *astigmatism* which becomes evident when rays of light pass through the lens obliquely and do not converge upon a common image point. This defect can be overcome by using two lenses suitably separated.



Chromatic aberration caused by different types of lenses

The high degree of accuracy required in the navigation of long-range guided missiles makes it obvious that any optical devices employed in missile navigation must be of the highest precision and free from the effects of undesired external influences.

Celestial navigation systems using fixed stars for references must be designed so that they are "sensitive" to only the selected stars and will not react to light rays of different intensity or color which might come within the fields of their telescopes. This means that all prisms used in the system must be accurately oriented with respect to the incident light reaching them through the "star-finding" telescope, so that the desired portion of the refracted spectrum is directed to the photo-cathode or photo-cell to maintain the standard reference voltage or current for the system.

All lenses so used must be free from chromatic aberration or carefully compensated by other lenses or color filters. They also must be free from spherical aberration and astigmatism since any minute deviation of the incident light beam within the optical system may produce a large "position error" in the navigation system.

RECTILINEAR PROPERTY OF LIGHT. Most astronomical and celestial navigation systems utilize the rectilinear property of light. Distances and angles are computed on the basis of *straight-line* measurements to the reference stars; therefore, any factor in the system which would introduce deviation or bending of light rays would cause an error in the straight-line computations. "Star-finding" telescopes are usually mounted so that the reference star, when seen along the principal axis of the lenses in the telescope, is also in a plane normal (perpendicular to the earth's atmospheric envelope. This positioning is maintained by means of a *stabilized-platform* mounting for the celestial navigation optical components and by accurate *parallel-sided* windows in the hull of the missile. The windows serve as a passageway through which the light from the celestial body must pass before reaching the tracking telescopes.

A *pin-hole camera* demonstrates the rectilinear propagation of light by producing a picture of an object on the film as a result of the photo-chemical action of rays of light from each point on the object. These rays pass through the pin-hole aperture in straight lines and cause chemical changes in the light-sensitive coating (emulsion) of the film or plate. The chemical changes in the coating are proportional to the amount of light received from corresponding points on the object.

The dimensions of the camera box and the size of the pin hole limit the size of the image. Enlarging the pin hole would increase the brightness of the picture and would permit the light rays to diverge from each point on the object and overlap on the film, thus resulting in a spotty or blurred picture.

The rectilinear characteristics of light and the principles of the pin-hole camera are illustrated on page 133.

These figures illustrate how enlarging the aperture of the pin-hole camera causes overlapping of light rays from the object due

to divergence of the rays from points on the object.

By placing a lens in the enlarged aperture of the camera, an additional brightness of the picture can be obtained without blurring or loss of definition. This is accomplished by adjusting the lens until the diverging rays from all points on the object are converged again at corresponding points on the film or plate in the camera. The bottom figure on the right illustrates how this is done.

Photographic cameras usually provide means for varying the distance between the lens and the film (focusing) and for adjusting the aperture of the lens to control the amount of light admitted and the exposure time required. These adjustments adapt the camera for use under a wide range of conditions of light and position.

Specially designed lenses, shutters, light filters, and synchronizing devices may be used to adapt the camera to specific applications and requirements, such as the photographing of objects moving at high rates of speed, distant objects, wide-angle views, underwater views, aerial maps, actions or operations synchronized with other events, and for preserving reference data.

The light-sensitive coating (emulsion) on the film or plate must also possess the proper characteristics for photographic applications under specific conditions. The photo-chemical action of the emulsion must be rapid for use in the photographing of fast-moving objects, since shutter speed must be fast to prevent blurring of the image.

You already are aware that telescopes utilize mirrors, prisms, and lenses. Now let's see how these optical devices function in the telescopes.

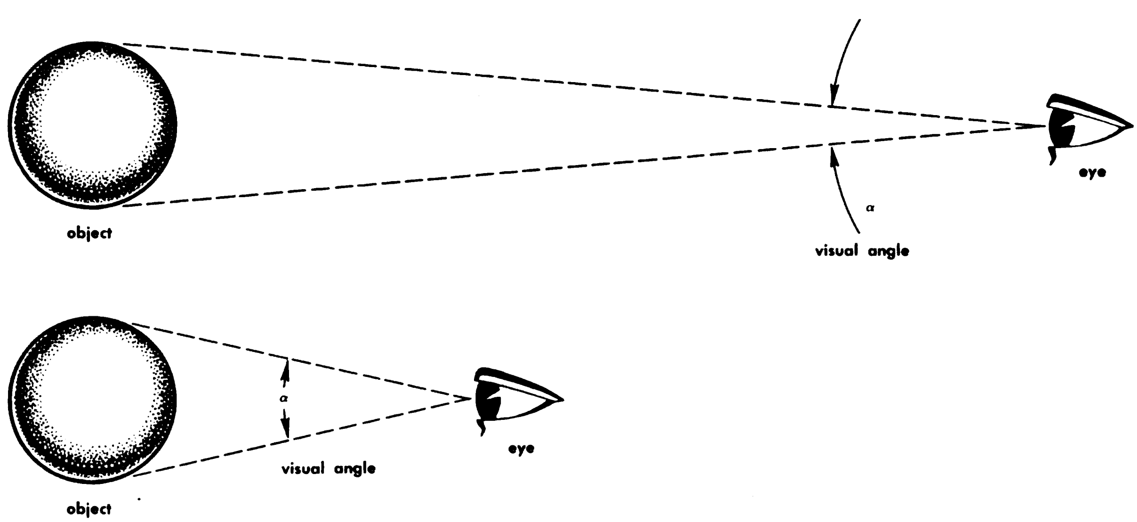
Telescopes

For our purposes, a telescope can be defined as an optical instrument which employs a combination of lenses or a combination of lenses and mirrors (or prisms) to magnify the apparent size of a distant object.

An object at a great distance from the eye will subtend a small visual angle, while the same object viewed at closer range will subtend a larger visual angle. This phenomenon is illustrated below.

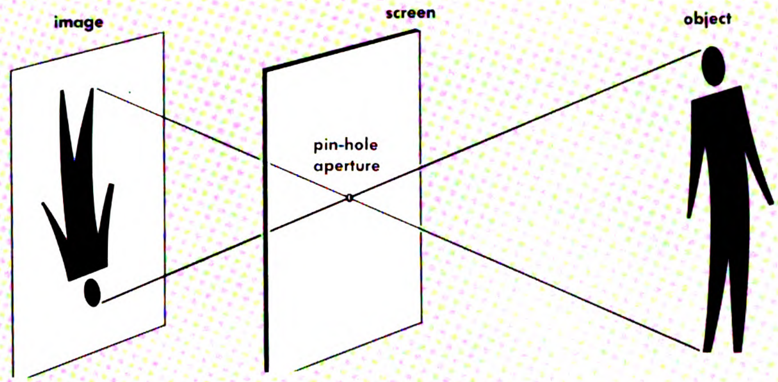
Magnifying power of an optical instrument may be defined as the ratio of the size of the image formed on the retina of the eye when the object is viewed with the aid of the instrument to the size of the image formed on the retina without the aid of the instrument.

The relative size of the images can be expressed in terms of the angles subtended by the images on the retina at a point along the optical axis of the eye and at the crystalline lens in the pupil of the eye.

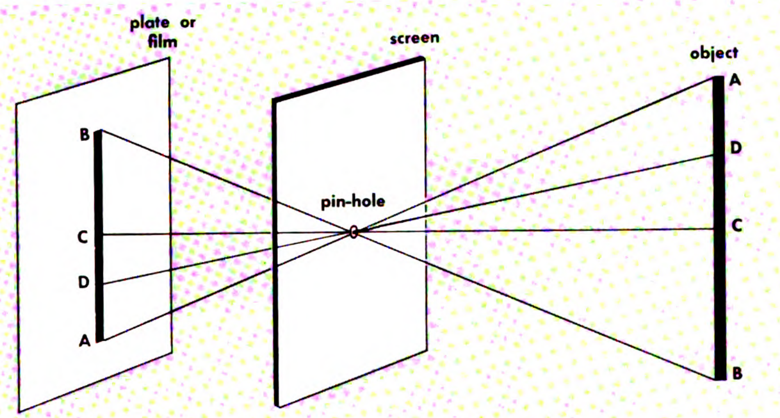


Relation between distance and visual angle

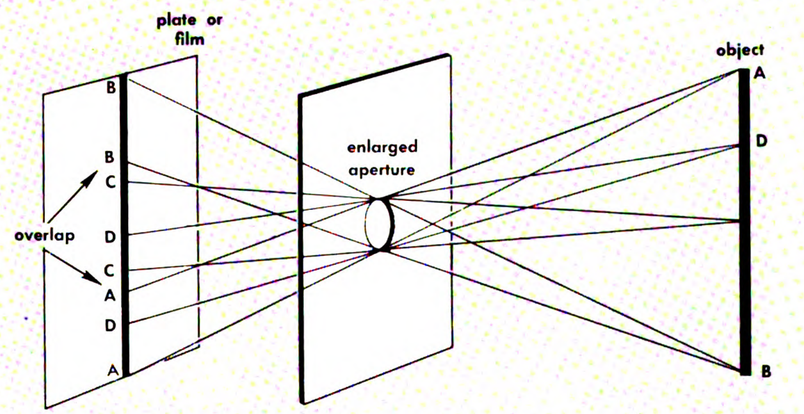
The rectilinear characteristics of light as demonstrated by the pin-hole camera.



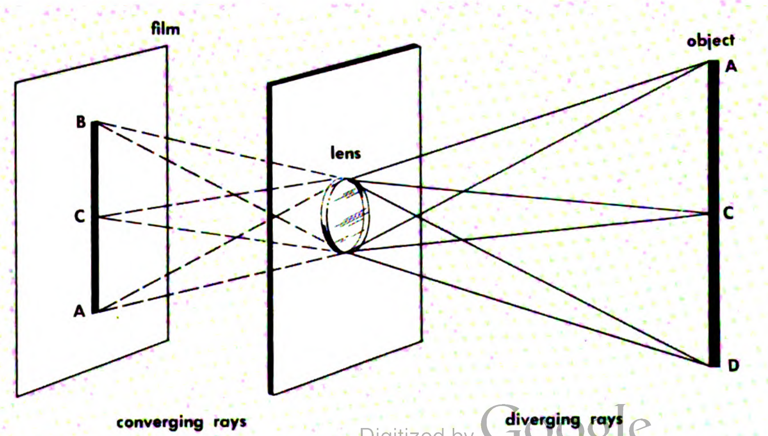
Camera with small aperture permitting no overlap of rays.



Camera with enlarged aperture resulting in overlapping of rays.



Action of lens in camera reconverges rays on film.



The size of the image on the retina depends upon the distance between the object and the eye; however, there is a limit to the size of the image that can be formed on the retina. By using a converging lens or magnifier, the image size on the retina is increased because the visual angle is increased as a result of a real image being formed closer to the eye than the object.

For high magnification, short-focal-length eyepieces must be used when objective forward lenses of long focal length are used. The short-focal-length eyepieces widen the angle of vision, thus rendering magnification.

Astronomical telescopes aid in viewing and photographing distant objects such as stars and planets by collecting and concentrating a larger beam of light, making the object visible or more distinct. There are many varieties of astronomical telescopes, but generally they are classified in two types: *refracting telescopes* and *reflecting telescopes*.

REFRACTING TELESCOPES. A refracting telescope is the type which employs two lenses or systems. One system includes the objective lens which serves to collect the beam of light and form the image. The image is then magnified by the other lens known as the eyepiece.

In this telescope, the objective lens has a focal length as long as possible (limited by the length of the telescope barrel) and produces a real image of the object. The eyepiece has a short focal length (10 inches or less) and produces a virtual image of the image produced by the objective lens. As mentioned before, the eyepiece magnifies the image by widening the angle of vision.

Optical arrangement of a refracting telescope is illustrated in the figure below.

The dotted lines represent rays from the upper end of a distant object. The rays converge at the lower end (point of arrow) of image I_A by the action of the objective lens.

These rays continue on and are deviated by the eyepiece so that they appear to come from the upper end of the inverted image I_B.

The solid lines represent the rays from the extremities of the object and image through the centers of the lenses.

Without the telescope, the distant object would produce an image on the retina subtending the angle of vision α, but with the telescope, the larger image subtending the angle β is produced on the retina.

The magnification of this telescope may be represented by the ratio of these two visual angles:

$$M = \frac{\beta}{\alpha}$$

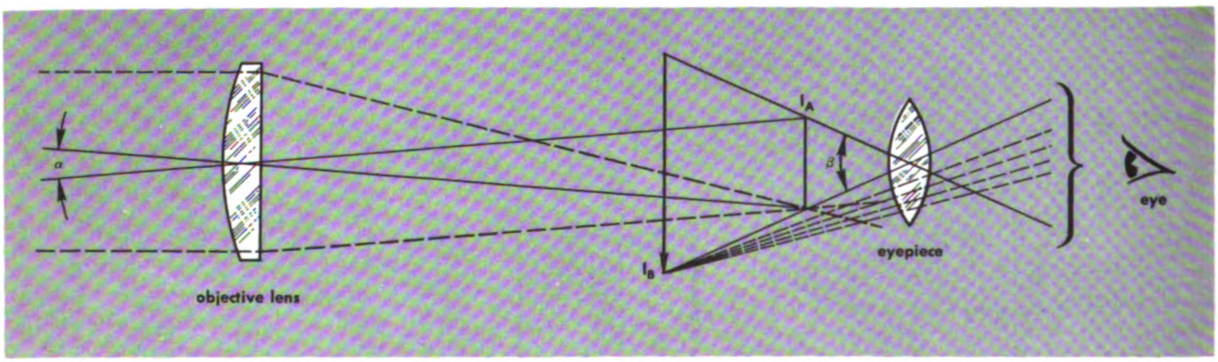
Since image I_A is located at the principal focus of the eyepiece which has a focal length "F" you will see that:

$$\alpha = \frac{I_A}{F} \text{ and } \beta = \frac{I_B}{F}$$

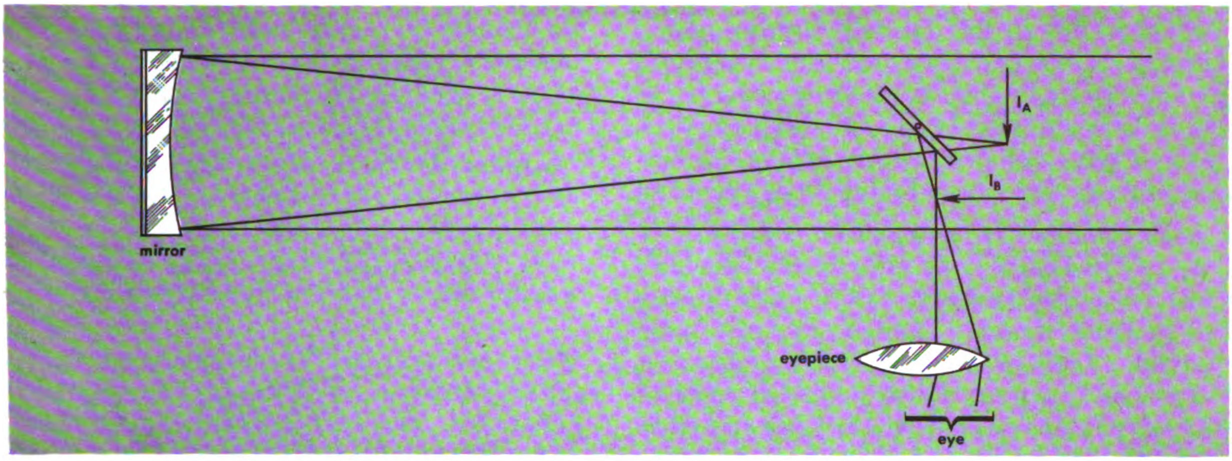
Thus, the magnifying power of the telescope (M) is equal to $\frac{F}{I}$.

The objective lens of the refracting telescope is a carefully corrected achromatic lens, and the eyepiece is generally a compound lens.

REFLECTING TELESCOPES. The second type of astronomical telescope is called a reflecting telescope. This type employs a concave mirror



Refracting astronomical telescope



Reflecting astronomical telescope

for collecting the beam of light and for forming the image. It also has an eyepiece for magnifying the image.

Reflector-type telescopes may be designed with shorter overall length and wider fields of vision than the refractor type of equal magnifying power. Consequently, most astronomical telescopes of high power are reflector-type instruments.

The figure above illustrates a reflecting telescope designed by Newton. In this telescope, a small plane mirror (or total-reflecting prism) on the axis of the telescope shifts the real image formed by the concave mirror from position I_A to position I_B. From I_B, the image is enlarged by the eyepiece which is set at right angles to the plane of the concave collecting mirror.

The same basic optical principles are in volved in both the refracting and reflecting-telescopes.

If a telescope is used for making photographic records, the eyepiece is removed and a sensitized plate or film is placed at the point where the image is formed by the objective lens. The optical arrangement, in this case, is the same as that of a camera with a great focal length.

In missile navigation, systems employing telescopes as star-trackers, refracting-type telescopes (with extremely accurate lenses of high magnification) and relatively narrow fields of view are used.

To this point we have discussed those optical devices in which only optical principles are involved. In the next few pages, the similarity of electromagnetic waves to light waves is pointed out. Television camera tubes, which employ electronic principles as well as optical principles also are discussed.

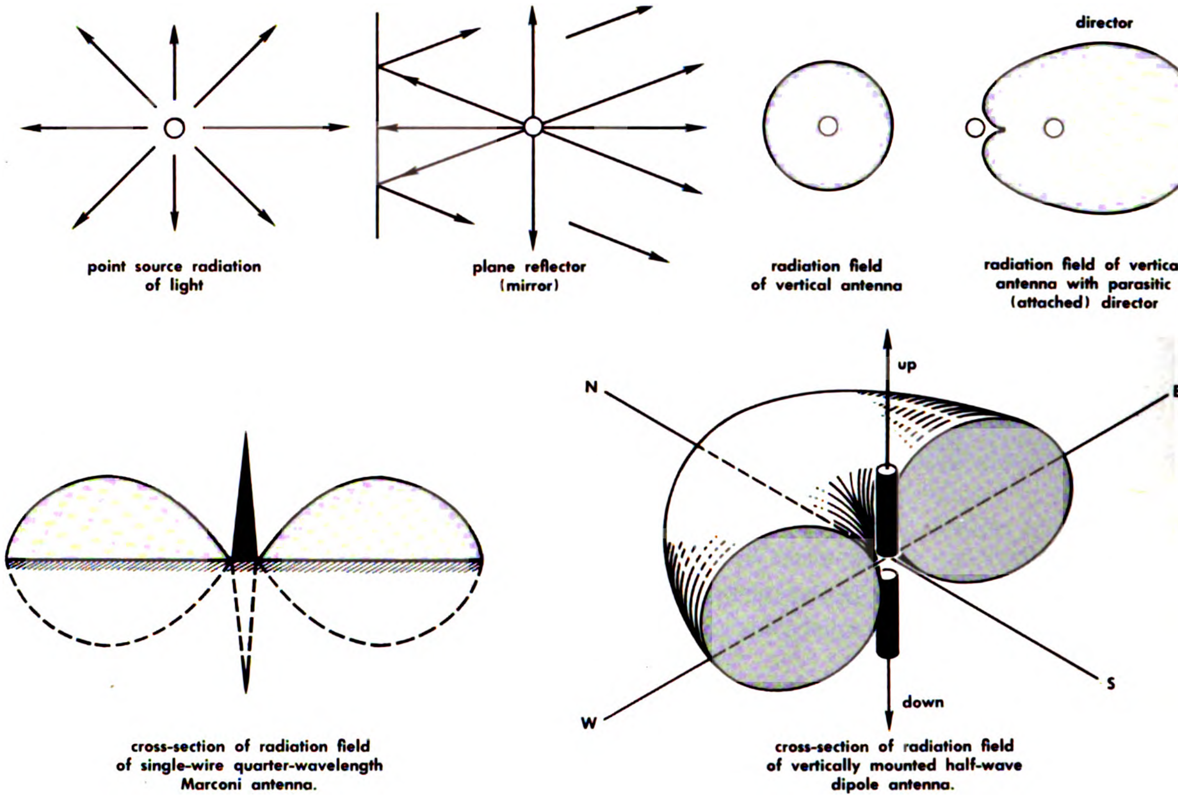
COMPARISON OF LIGHT AND ELECTROMAGNETIC WAVES

To illustrate the similarity between light and electromagnetic waves, you attention is directed to a comparison of optical systems and directional and beam-forming antenna systems, such as are employed with radio direction finders and radar equipment.

Radiation of light from a point source in space is omnidirectional; that is, the light rays emanate in all directions along straight-line paths. By means of reflectors, lenses, and filters, light rays can be directed, diffused, concentrated, selected as to wavelength (color), and focused upon specific points as desired.

Radiation of electromagnetic waves from a single-wire vertical antenna (Marconi type) or from a vertically mounted dipole is in the form of concentric fields along the horizontal plane as illustrated in the figures on the next page.

By means of suitable reflector and director elements, the electromagnetic radiations are directed and concentrated (beamed) to produce a radiation pattern of the desired form.



Antenna radiation of electromagnetic waves and radiation of light waves

Also by *phasing*, or feeding two or more antenna elements in varying degrees of phase with respect to each other, the fields produced around each element can be made to reinforce or to cancel each other. The reinforcing or canceling is done in such a manner that the individual fields combine vectorially to produce a radiation field of maximum intensity along the desired paths in either or both horizontal and vertical planes.

Where a narrow and intense beam of energy is required, as in tracking radar applications, parabolic reflectors or combinations of reflector and director elements may be used to form the beam. The beam is formed in much the same manner that the reflector and lens in a "bulls-eye" lamp or focusing flashlight shape the beam of light. Basically, any antenna is a conductor or system of conductors used to radiate or receive energy in the form of electromagnetic waves. The height of the radiating element above ground, the conductivity of the earth below it, the wavelength of the radio propagation, and the shape and dimensions of the antenna all affect the pattern of the field radiated into space. Thus, by controlling these factors, it is possible to produce the desired radiation pattern.

If you should want to direct the radiation from a source of light in one general direction, you could place a reflecting surface behind the source in such a manner as to direct all the rays toward the object to be illuminated. The rays which normally would follow paths in directions away from the object to be illuminated are reflected back toward the object and reinforce the rays normally traveling toward it.

If the object is small or if you desire maximum intensity of illumination on a specific point of the object, you could cause a greater number of the omni-directional rays to fall upon the desired point by using a curved reflector which would cause the light rays to

converge at the desired point. A reflector of specific curvature located at a fixed distance from the source causes the light rays to converge at a given point along its principal axis.

However, if it should be undesirable or impractical to vary either the source or the reflector, you could employ a movable lens between the object and the source to focus the reflected beams, as illustrated in the figures below.

The foregoing effects are obtained with either light or high-frequency electromagnetic waves by the use of parabolic reflectors. Parasitic (attached) reflectors and directors may be used to cause divergence or convergence of electromagnetic waves, thus serving as forms of electron lenses.

Electronic and Optical Principles of Camera Tubes

A good example of a device that employs both optical and electronic principles is a television-camera tube, used to pick up the scene to be transmitted. Several types of camera tubes are in use, and while they vary in mechanical details, their principles of operation are basically the same. Here we will consider the "Iconoscope," a camera tube widely used today.

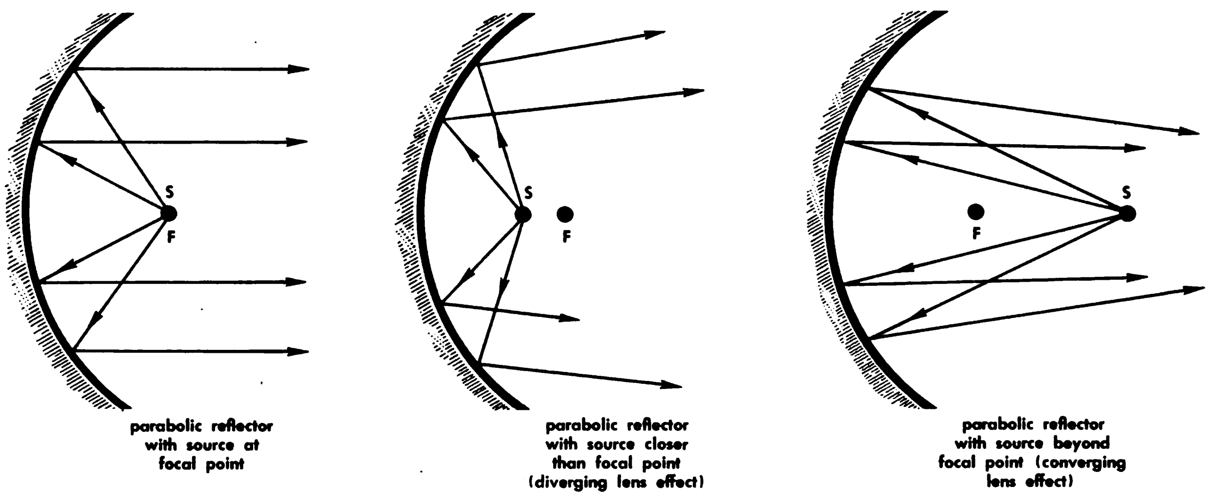
The "Iconoscope" consists of a glass envelope shaped somewhat like a saucepan with an enlarged handle. The side of the envelope corresponding to the top of the pan is a transparent window through which the image of the object is permitted to fall upon a screen or *mosaic*, composed of thousands of photosensitive globules. An electron gun similar to that of the common cathode-ray tube is used to project a beam of electrons over the mosaic.

This beam, known as a scanning beam, removes the charge from each section or globule of the mosaic in a fixed sequence. And in so doing, it produces discharge currents in proportion to the quantity of light received upon each section of the mosaic. These currents serve to amplitude modulate the transmitter and develop the picture intelligence in the television system.

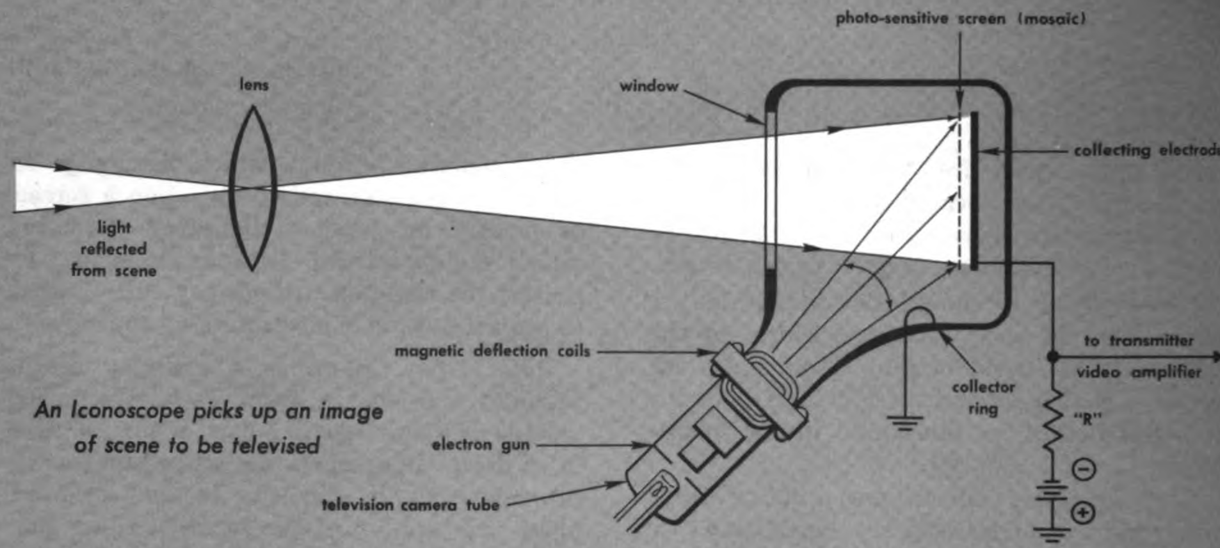
The illustration on the next page of the "Iconoscope" shows some of the associated components used to convert variations in intensity of the light, reflected from the object being televised, into variable currents.

The object or scene being televised is focused on the photosensitive mosaic which contains the thousands of minute globules. The globules are electrically insulated from each other.

The photosensitive screen consists of a thin sheet of mica. On one side of the mica sheet are the globules made up of a light-sensitive compound of silver and cesium. On the reverse side of the mica sheet is a thin metal coating that serves as a collecting electrode. Each light-sensitive globule is effectively one elec-



A reflected beam can be focused by employing a movable lens



An Iconoscope picks up an image of scene to be televised

trode of a tiny capacitor, while the metal coating on the back of the mica sheet serves as a common electrode for all the globules. Thus, the mosaic comprises many tiny capacitors all having one common electrode, and when the optical image is focused upon it, the reflected rays strike the tiny photocells causing them to lose electrons through secondary emission. As stated earlier the number of electrons lost by each tiny element is proportional to the intensity of the light ray which strikes it. These released electrons are drawn to a *collector ring* (graphite coating) on the inner surface of the tube. The globules from which they were released now have a positive charge with respect to the back plate or collecting electrode of the mosaic.

The portions of the mosaic which are most brilliantly illuminated emit the greatest number of electrons and therefore have the highest positive charge.

After exposure, the mosaic is scanned sectionally in a regular sequence by a magnetically deflected beam of electrons from the electron gun. This scanning beam neutralizes each element of the mosaic and causes current to flow through the collector load resistor ("R" in the illustration) in an amount proportionate to the positive charge existing on each element struck by the scanning beam.

The voltage pulse developed across the resistor represents the video signal and is amplified to a suitable level for modulation of the transmitter's carrier wave. Thus, the picture intelligence is transmitted to the receiver in the form of an amplitude-modulated carrier.

The receiver employs a conventional amplifying and demodulation system to separate the intelligence from the carrier. This separation is in the form of a varying voltage which is applied to the signal grid (control grid) of the cathode-ray-type picture tube or kinescope, on the screen of which the image is reproduced.

Thus, a scene is transformed from variations in light to variations in current and back to variations of light again by means of combined optical and electronic processes. In this manner, a scene may be transmitted through a considerable distance.

AN AID TO UNDERSTANDING MISSILE SYSTEMS

As indicated earlier, the intent in this section was to point out the similarity between the characteristics of light and those of radio propagation and to show how electromagnetic and light waves can function together in electronic equipment.

An optical system controls radiant energy by means of mirrors, reflectors, prisms, and lenses so that the radiant energy may be converted into a form suitable to the requirement of the application; an antenna system also controls electromagnetic radiations in part by means of reflectors and lenses (directors). When you study the control and navigation systems of missiles, you will find this information helpful toward a clear understanding of the systems.

Physics of Transistors

Transistors are crystal substitutes for the conventional vacuum tube. They show promise of replacing vacuum tubes in many types of airborne electronic units. The advantages of using transistors will become clear to you as you read through this section.

Since the early days of the radio era when G. W. Pickard first discovered the rectifying properties of certain crystals, extensive research has led to an ever-increasing use of crystals in electronic applications. Crystals are especially applicable in devices in which space and weight are essential factors.

Quartz crystals are widely used for controlling the frequency of oscillators; Rochelle salts are used in piezo-electric devices such as phonograph pickups and microphones; galena and silicon crystals are widely used in VHF and UHF radar as detectors and mixers. Each of these crystals possesses some characteristic which makes it suitable for one specific application. The most versatile crystal yet found is a *germanium* crystal.

Only in recent years has the versatility of germanium for electronic purposes been known. The germanium crystal has since become prominent in the field of electronics, resulting in the development of a family of crystal triodes, tetrodes, and pentodes, known as transistors.

STRUCTURE OF CRYSTALS AND ATOMS

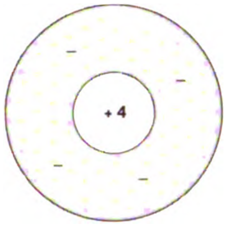
To understand the theory of the transistor, you will find it helpful to consider briefly the structures of atoms and crystals.

As you know, all matter is composed of one or more elements, and each element is composed of atoms. The atom, in turn, is composed of smaller units of matter called electrons, protons, and neutrons. Electrons possess a negative charge. Protons possess an equivalent positive charge and are more than 1800 times as heavy as the electron. The neutron possesses no charge and has the same mass as the proton.

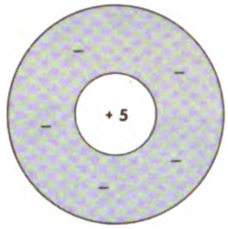
The atom, which is the smallest subdivision of an element possessing all the properties of the element, is composed of a nucleus or core of protons and neutrons around which the lighter electrons revolve. In the normal atom there are as many electrons (negative charges) outside the nucleus as there are protons (positive charges) within the nucleus. Thus, the atom is balanced, or electrically neutral.

Atoms of one chemical element differ from those of another element only in the number of electrons, protons, and neutrons in their structure. With reference to transistors, we will consider only a few elements. These elements are germanium, silicon, antimony, arsenic, aluminum, gallium, indium, and thallium.

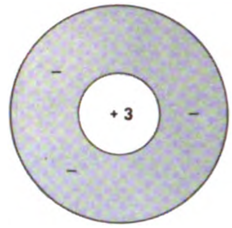
The study of atomic structure has shown that a large portion of the electrons around the nucleus are tightly bound to it and do not enter into chemical reactions or transistor physics. The nucleus and the tightly bound electrons comprise an inert core with a net positive charge around which the less tightly bound electrons revolve.



Atomic structures of germanium or silicon



Atomic structures of antimony or arsenic (donors)



Atomic structures of aluminum, gallium, indium or thallium (acceptors)

In transistor physics, we'll be concerned with the net charge on the core and the electrons surrounding the nucleus. For example, each atom of germanium has 32 protons in its nucleus and 28 tightly bound electrons around it; thus the atoms of germanium can be represented by a core with a net charge of +4 surrounded by four electrons. Each atom of silicon, which has 14 protons in its nucleus and ten tightly bound electrons around it, is represented in the same manner as an atom of germanium, as shown in the left figure above.

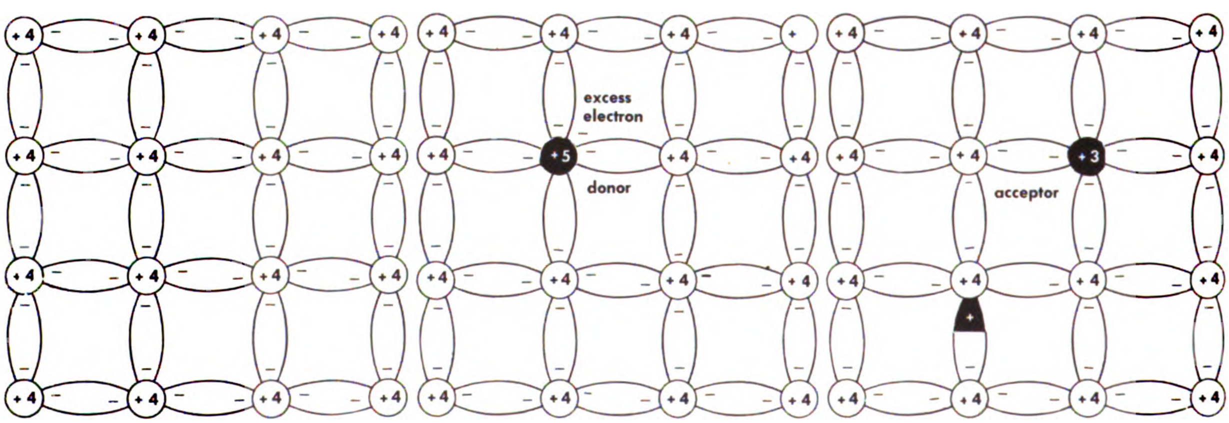
Pure germanium can not be used in producing transistors because germanium acquires the properties of rectification and amplification of current through the presence of impurities in the crystals. One type of impurity is known as *donor* and another type is called *acceptor*. They have these names because of the manner in which they affect the electron movement within the transistor.

Antimony and arsenic become *donors* when they join the crystal structure of germanium. The net charge on the cores of antimony and arsenic atoms is +5. Each atom has five electrons surrounding its core, as shown in the center figure above.

Aluminum, gallium, indium, and thallium become *acceptors* when they join the crystal structure of germanium. The net charge on the cores of their atoms is +3, and each of their atoms has three electrons surrounding its core. The atomic structure of these acceptors is pictured to the right above.

The figures below compare the atomic conditions existing in a pure germanium crystal with those existing in crystals containing donor and acceptor impurities.

In the pure crystal each atom has four neighbors which are equidistant from each other. Between the cores of the atoms and each of their neighbors are two electrons (shaded area). These paired electrons form



Atomic conditions in germanium crystals

electron-pair bonds which come into existence when two or more atoms approach each other. Since the electrons are in constant motion around the core, these electron-pair bonds are formed when the movement of an electron from one atom becomes coordinated with the movement of an electron from another atom.

This coordination tends to attract the cores toward each other, but the cores' positive charges repel each other until they attain a perfect balance of attraction and repulsion. The atoms are then said to be in a condition of equilibrium.

As previously mentioned, pure germanium crystals can not be used as transistors. This condition exists because the atoms of the crystals would be in a state of equilibrium, in which form they constitute good insulators with a dielectric constant of approximately 16.

Effects of Donor Atoms on Germanium Crystals

When a donor atom (antimony or arsenic) joins the crystal structure of germanium, the atom must lose one of the five electrons surrounding its core because only four of its electrons can form electron-pair bonds with the electrons of its neighboring germanium atoms. Therefore, the excess electron is free to move through the relatively wide spaces between the cores. It moves through the crystal as though it were in a vacuum.

If a battery were placed across the crystal, the electron would move toward its positive terminal and enter the battery at that point. Simultaneously, an electron would leave the negative terminal of the battery and enter the crystal. A continuous electron flow through the crystal would take place, but the cores of the germanium and donor atoms would remain undistrubed.

Germanium crystals containing donor impurities are known as N-type germanium. You will note that the designation "N" results from the fact that conduction through the crystal is mainly a conduction of *negative* charges in the form of excess electrons from the donor atoms. It is this action which led to the term *donor* being applied to the element, which when joined to the crystal structure of germanium gives off an excess electron.

Effects of Acceptor Atoms on Germanium Crystals

When an acceptor atom (aluminum, gallium, indium, or thallium) joins the crystal structure of germanium, it does so by accepting an electron from one of its neighboring germanium atoms. Atoms of acceptors have three electrons surrounding each of their cores, while the cores of germanium atoms have four electrons surrounding them. When this new electron-pair bond is formed, a *hole* is created in another electron-pair bond. This *hole* possesses the equivalent charge of an electron, but it is a positive charge.

Experiments have shown that the *hole* is free to move throughout the crystral structure; thus, conduction of current can take place through a germanium crystal containing an acceptor as readily as through one containing a donor. However, the conduction process is different.

If a battery were placed across this crystal, the hole would be attracted toward the negative terminal of the battery and an electron from this terminal would enter the crystal and fill the hole.

Simultaneously, an electron from one of the electron-pair bonds in the crystal and near the positive terminal of the battery would separate from its bond and enter the battery, thus creating another hole in the crystal. This action would repeat and maintain a continuous flow of current through the crystal.

Germanium crystals containing acceptor impurities are known as P-type germanium, the term being derived from the fact that in such crystals the conduction through the crystal is mainly a conduction of positive charges (holes).

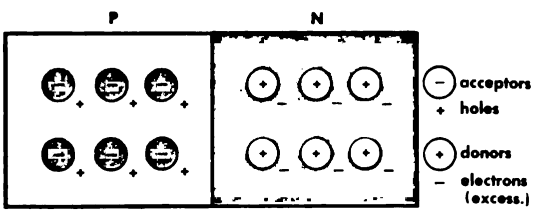
A hole can be defined as an incomplete group of electrons whose general properties are similar to those of an electron except that it carries a positive charge instead of a negative one.

An acceptor is defined as an element which, when it joins the crystal structure of germanium, produces a hole or excess positive charge within the crystal.

P-N TYPE JUNCTION

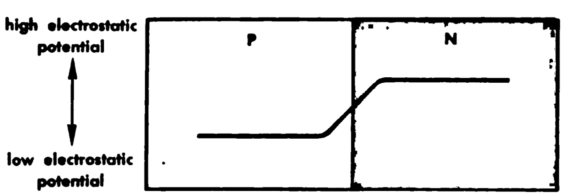
You must bear in mind that germanium crystals are good conductors and can conduct current equally well in either direction. Rectification with a germanium crystal occurs if P-type germanium and N-type germanium are placed side by side. The plane at which the two types of germanium meet is called a P-N junction, and the action which occurs at the junctions constitutes the basic action of transistor operation.

The figure below illustrates a P-N-type junction in a state of equilibrium. Notice that the holes concentrate to the right of the P-type germanium atoms and the excess electrons concentrate to the right of the



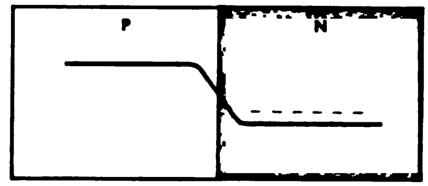
P-N junction in condition of equilibrium

N-type germanium atoms. This phenomenon is due to the distribution of electrostatic potential produced by the acceptor and donor atoms as illustrated below.



Electrostatic potential

The electrons remain in the region of *highest electrostatic potential* and the holes remain in the region of *lowest electrostatic potential.* When an electron is in the region of highest electrostatic potential, its *potential energy* is at a minimum as shown next. Since *potential energy* means ability to do work, and since the electron cannot move to do work after it reaches the point of highest electrostatic

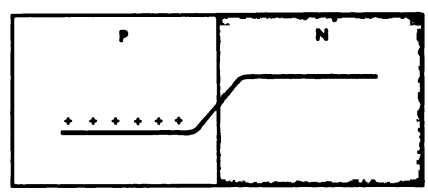


Potential energy of electrons

potential, it becomes apparent that the electron is in a region of low potential energy.

Bear the preceding statement in mind, because potential energy diagrams are used often in illustrating transistor behavior.

The same statement can be applied with reference to holes, for when the hole is in a region of lowest electrostatic potential (that is, low negative potential), its potential energy is at a minimum as illustrated below.

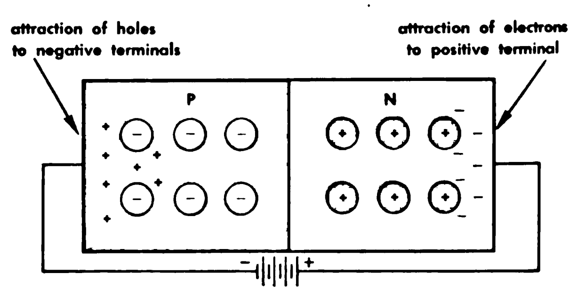


Potential energy of holes

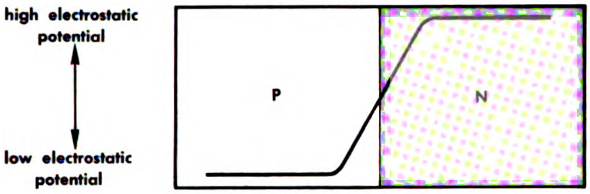
Holes and electrons flow into regions of low potential energy only. A low-potential-energy region for an electron is a high-potential-energy region for a hole, and vice versa.

P-N Junction with Reverse Bias

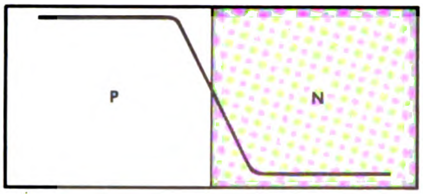
If you connected a battery across a P-N germanium crystal as shown in the figure below, there would be no conduction of current through the crystal. Such a connection is called reverse bias.



P-N junction with reverse bias



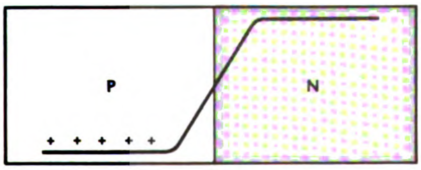
Electrostatic potential increased in P-N junction with reverse bias



Potential energy of electrons
The hill has been increased due to reverse bias

This condition occurs when the positive terminal of the battery is connected to the N-type germanium and the negative terminal is connected to the P-type germanium.

The positive terminal of the battery attracts the electrons and causes them to concentrate farther to the right than when the junction is in a condition of equilibrium. The negative terminal of the battery attracts the holes and causes them to concentrate farther to the left than when the junction is in a condition of equilibrium. As a result of these attractions, there is no flow of electrons to the left and no flow of holes to the right. However, the difference in electrostatic potential between the two types of germanium has been increased as shown in the above figure.



Potential energy of holes
The hill has been increased due to reverse bias

The *potential-energy hill* for the holes has been increased, thus, they will not flow up the steep hill. This condition is shown above.

The same condition applies to the *potential-energy hill* for the electrons and thus prevents their flow.

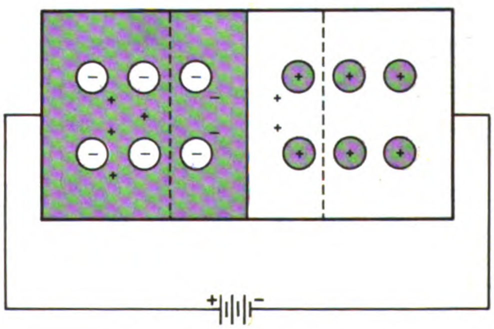
P-N Junction with Forward Bias

If you connected a battery across a P-N germanium crystal with the positive terminal connected to the P-type germanium and the negative terminal to the N-type germanium, current would flow in proportion to the applied voltage. This type of connection is known as *forward bias.*

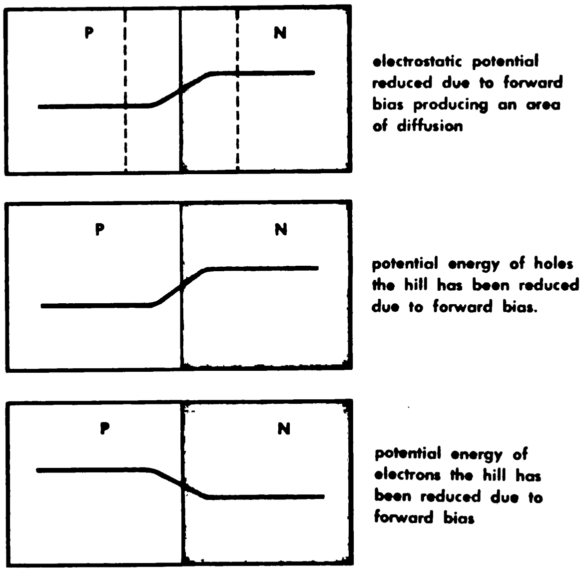
As shown in the illustration below, the positive terminal of the battery repels the holes and causes them to move toward N-type germanium. Some of the holes enter the "N" area.

The negative terminal of the battery repels the electrons and causes them to move toward the P-type germanium. And, in this case, some electrons enter the "P" area.

Electrons and holes combine in a small area of diffusion on either side of the P-N junction (between the dotted lines in the diagram). For each hole in the "P" region that combines with an electron from the "N" region, an electron from an electron-pair bond in the crystal near the positive terminal of the battery enters the battery at the positive terminal. This action creates a new hole which moves toward the N-type germanium. For each electron that combines with a hole in the N-type germanium, an electron enters the crystal from the negative terminal of the battery. The current flow in the "P" region is mainly a flow of holes, while that in the "N" region is mainly a flow of electrons.



P-N junction with forward bias

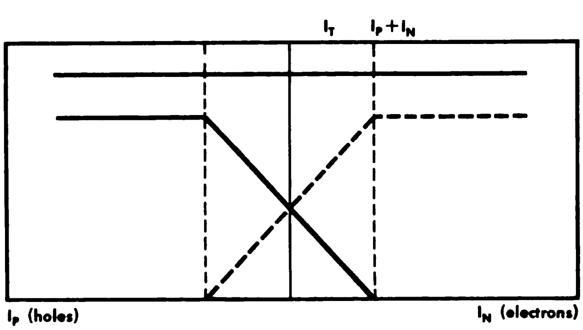


Potential energy conditions existing under application of forward bias

The potential-energy conditions existing under the application of the forward bias are shown in the accompanying figures.

The currents that flow through the crystal are shown below. The total current (I_T) is constant; current by holes (I_P) is shown by the solid line; and current by electrons (I_N) is shown by the dashed line.

From the above theory, you can understand more fully the process of rectification in germanium crystals; you can better understand how the crystal acts as a low resistance when forward bias (+ to P, − to N) is applied, and how it acts as a high resistance when reverse bias (− to P, + to N) is applied. These phenomena constitute the basic principle of transistor operation.



Current flow through crystal under application of forward bias

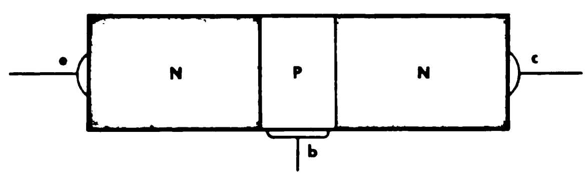
Transistors, as presently developed, fall into two general classification: junction transistors and point-contact transistors. Junction transistors will be discussed first.

N-P-N AND P-N-P JUNCTION TRANSISTORS

In this section we will duscuss N-P-N and P-N-P junction transistors. The N-P-N transistor is constructed by placing a narrow strip of P-type germanium between two relatively long strips of N-type germanium. And, as the letters indicate, the P-N-P transistor consists of a narrow strip of N-type germanium between two relatively long strips of P-type germanium.

N-P-N Junction Transistors

An N-P-N junction transistor is pictured at the right. In this type, large surface contacts (low-resistance) are made to each strip. The N-type germanium, as shown at the left, is called the emitter (e). The N-type strip, shown at the right, is called the collector (c). The P-type section in the center is called the base (b).



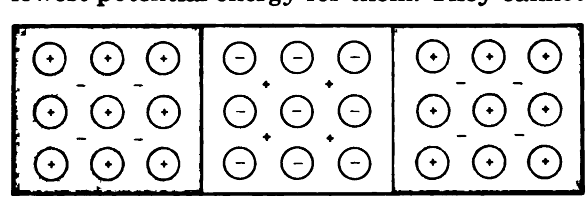
N-P-N Junction Transistor

The illustration below shows the distribution of the donors, acceptors, holes, and electrons in an N-P-N junction transistor under equilibrium conditions (no external voltages applied).

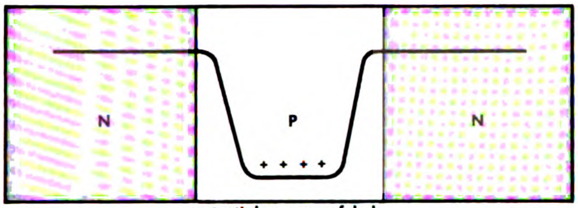
The potential energy in this type of transistor is represented on the right.

The second figure on the next page shows the potential energy region of the electrons.

The holes are concentrated in the region of lowest potential energy for them. They cannot

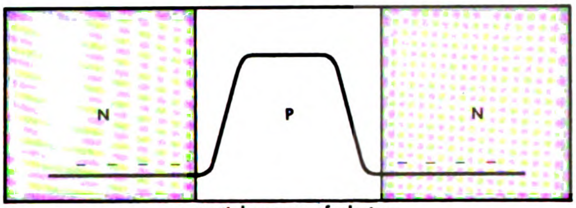


N-P-N junction transistor in state of equilibrium



potential energy of hole

Potential energy of hole in N-P-N junction transistor



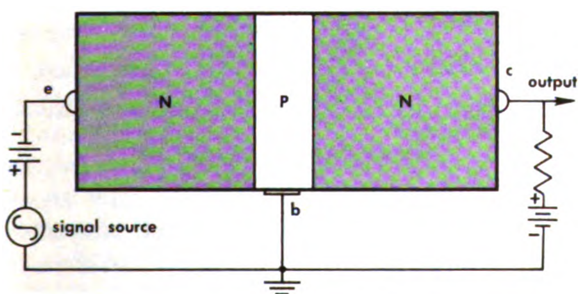
potential energy of electron

Potential energy of electrons in N-P-N junction transistor

climb the potential-energy hills to the right or left.

The electrons are concentrated in the region of lowest potential energy for them and cannot climb the potential-energy hills to enter the P-type germanium. Thus, no current flows.

In practical applications, the N-P-N junction transistor normally is biased in the manner shown in the accompanying diagram. The P-N junction between the emitter (e) and the base (b) is biased in the forward direction.



N-P-N junction transistor operating circuit

The P-N junction between the collector (c) and the base (b) is biased in the reverse direction.

The figure at upper right shows the potential energy condition for electrons, with no signal applied. In N-P-N transistors, the electrons are the major current components, so potential-energy diagrams for holes are not shown.

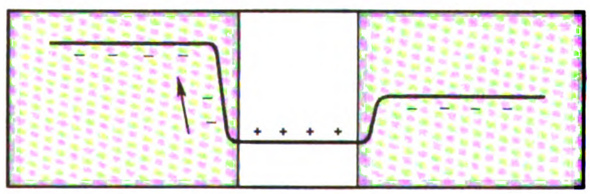
Potential energy of an electron (no signal)

The forward bias applied between the emitter and the base reduces the potential-energy hill at the left-hand P-N junction, so some electrons climb this hill and enter the P-type germanium. Since the base strip is relatively thin, most of the electrons which enter it will not combine with holes; instead they pass through the strip and readily go down the potential-energy slope at the right P-N junction, as shown by the arrow in the preceding figure.

The steep potential-energy slope which permits easy entrance of the electrons from the base strip to the N-type germanium of the collector is produced by the application of reverse bias between the collector and the base.

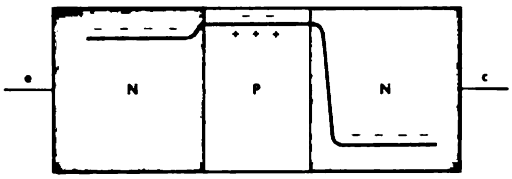
When a signal is applied which opposes the forward bias on the emitter (makes base more negative with respect to emitter), the potential-energy hill between the emitter and the base is increased and fewer electrons climb the hill to enter the P-type germanium.

The electrons which do enter the P-type germanium do not recombine with holes but fall into the collector region of low potential energy.



Signal applied to N-P-N junction transistor making base more negative with respect to emitter

When the polarity of the applied signal aids the forward bias (makes base more positive with respect to emitter), the potential-energy hill between the emitter and the base is decreased and more electrons flow into the P-type region. Most of these electrons do not



Signal applied to N-P-N junction transistors making base more positive with respect to emitter

combine with holes but flow readily to the low potential-energy level of the N-type germanium of the collector. This condition is demonstrated in the accompanying illustration.

You can compare the operation of the N-P-N junction transistor to the operation of the triode vacuum tube. The emitter is equivalent to the cathode, the base to the grid, and the collector to the plate.

Practically all of the electrons which emerge from the emitter (cathode) go to the collector (plate). The base (grid) current is minute and consists only of a small number of electrons which represent the recombination number of holes and electrons in the P-type region.

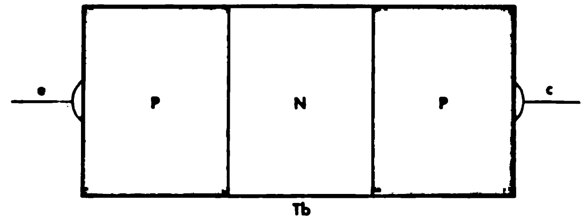
Two types of current are present in the vacuum tube, and two types of current are involved in the transistor. The thermionic electrons (those emitted from the cathode) of the tube are equivalent to the excess electrons from the emitter.

The minute conduction current of electrons (those which flow in and out of the grid to control plate current) is equivalent to the small current of holes in the base region (which varies to control collector current).

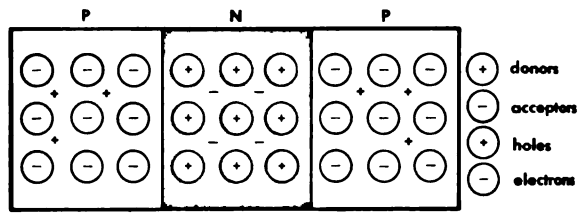
P-N-P Junction Transistors

As stated before, the P-N-P junction transistor consists of a narrow strip of N-type germanium between two relatively long strips of P-type germanium. Large surface contacts (low-resistance) are made to each strip. The contacts are designated b, c, and e, representing the base, collector, and emitter, respectively, as in the case of the N-P-N transistor.

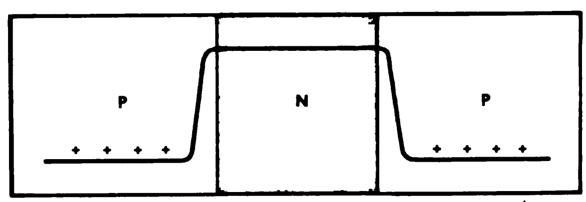
The operation of the P-N-P junction transistor is analogous to that of the N-P-N, except that the hole constitutes the main current component instead of the electron.



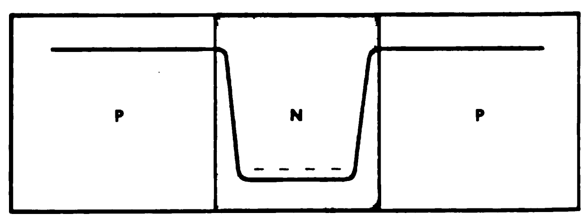
P-N-P junction transistor



P-N-P junction transistor in state of equilibrium



Potential energy of hole

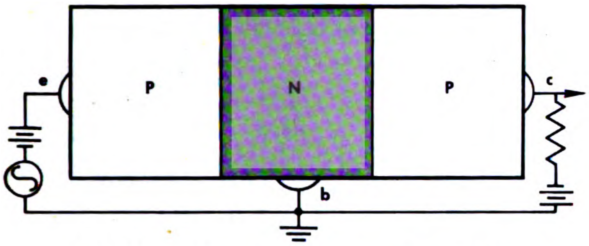


Potential energy of electron

Pictured on the right is the P-N-P transistor in a state of equilibrium.

The figures on the next page illustrate the conditions of operation for the P-N-P junction transistor in the same manner that the earlier figures illustrated the conditions of the N-P-N transistor.

Note that in order to bias the P-N junction between the emitter and the base of the P-N-P transistor in the forward direction, the emitter must be made positive with respect to the base; and to bias the collector in the reverse direction, the collector must be made negative with respect to the base.



P-N-P junction transistor operating circuit

The excess electron is the key factor in the operation of the P-N-P junction transistor, but the hole is the main current component. Basic conditions determining the potential energy of the hole are shown by the accompanying figures.

As you now know, the emitter, base, and collector connections to the crystal are low-resistance (large-area) connections in a junction transistor. In this respect, junction transistors differ from point-contact transistors, except with regard to the base connection.

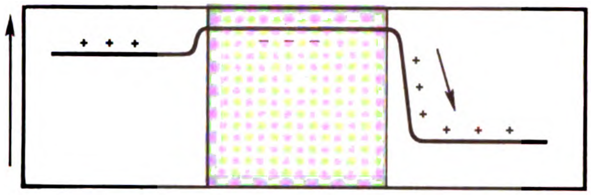
POINT-CONTACT TRANSISTORS

In a point-contact transistor, the emitter and collector connections to the crystal are point-contact (relatively high-resistance, small-area) connections. The adjoining figure illustrates a cut-away view of this type of transistor. The base connection, it will be noted, is a large-area, low-resistance connection.

As already mentioned, the emitter current is slightly larger than the collector current in a *junction transistor*, and the base current is extremely small. It has been found by experiment that in a *point-contact transistor*, the collector current (unlike that of the junction transistor) is substantially larger than the emitter current and that the base current is relatively large.

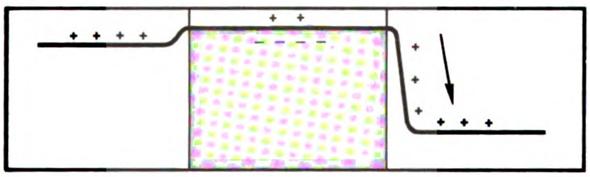
These differences between the junction transistor and the point-contact transistor are treated in succeeding paragraphs.

Although the point-contact transistor employs what is considered to be either an N-type germanium crystal or a P-type germanium only, experimental investigation has shown that in the N-type point-contact transistor, P-type layers occur; and that in the P-type point-contact transistors, N-type layers occur.

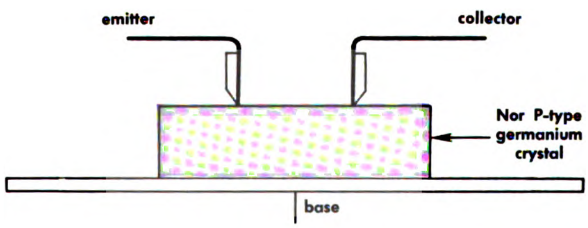


P-N-P transistor (no signal)

P-N-P transistor signal applied, making base more positive with respect to emitter



P-N-P transistor signal applied making base more negative with respect to emitter

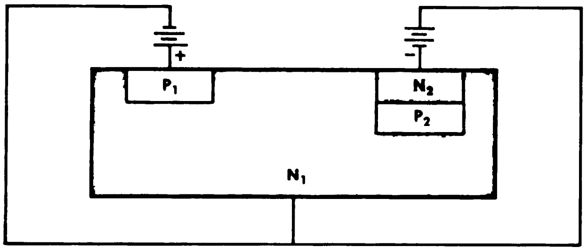


Point-contact transistor

N-Type Point-Contact Transistors

The illustration of an N-type point-contact transistor, on the next page, shows that under the emitter point there is a thin layer of N-type germanium (N_2) followed by a thin layer of P-type germanium (P_2).

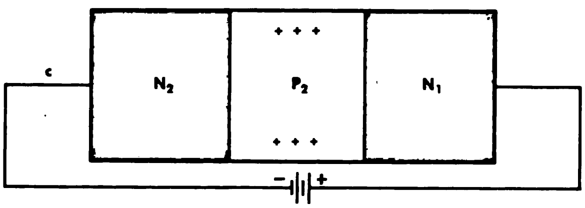
You can see in this drawing that the emitter-base section is biased in the forward direction. Current flow, consisting mainly of holes in the P_1-type germanium and of electrons in the N_1-type germanium, will occur. Under these conditions, an increase in emitter current causes a large increase in collector current, and a decrease in emitter



Construction of an N-type point-contact transistor

current causes a large decrease in collector current.

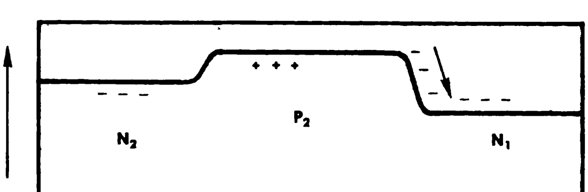
CURRENT AMPLIFICATION. To help you understand the current amplification of a point-contact transistor, we'll consider the collector-base circuit of the transistor. An enlarged view of the collector-base circuit appears below.



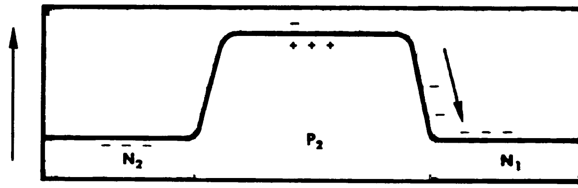
Collector-base circuit of an N-type point-contact transistor

The potential energy of electrons in the crystal portion of this circuit is shown in the diagram below. Under steady-state conditions (no signal applied), few electrons climb the potential-energy hill between N_2 and P_2. Those that do climb the hill do not combine with the hole in the P_2 region because it is a thin region. Instead, these electrons fall quickly down the potential hill between P_2 and N_1.

When the emitter-base region conducts, some of the holes which leave the P_1 region of the emitter drift into the P_2 region of the collector. With holes in the P_2 region from the



Potential energy of electrons without holes from emitter



Potential energy of electrons with holes from emitter

emitter, the potential-energy diagram for electrons between the collector and base conforms to that shown in the preceding figure.

Note that the potential-energy hill between N_2 and P_2 has been substantially reduced, and electrons in the N_2 region can climb readily into the P_2 region.

Most of these electrons, instead of combining with holes in this region, fall rapidly into the N_2 base region. For every hole that enters the P_2 region from the P_1 region, many electrons flow from the N_2 region through the P_2 region down into the N_1 region. Thus, it is evident that large amplification of current results.

This amplification of current is expressed as α (alpha) and is equivalent to the voltage amplification factor of a vacuum tube (μ). Mathematically,

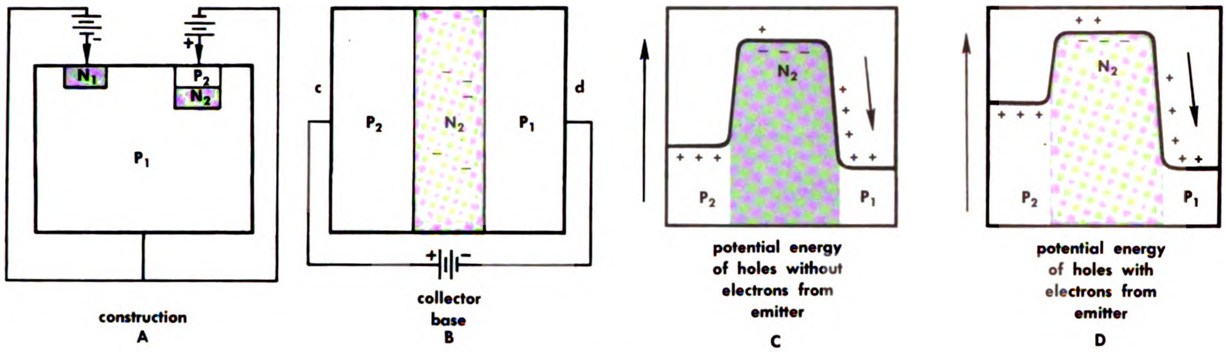
$$\alpha = \frac{\Delta I_c}{\Delta I_e}$$

where ΔI_c is a small change in collector current, and ΔI_e a small change in emitter current.

P-Type Point-Contact Transistors

Point-contact transistors have been made with P-type germanium as the main body. The theory explaining the operation of this type of transistor is similar to that explaining the operation of the N-type. The difference is that the main current-carrier consists of holes instead of electrons.

The explanation in the preceding paragraphs coupled with the data in the following figures will enable the reader to understand the theory of operation of the P-type point-contact transistors. Note that the polarities for the emitter with respect to the base, and of the collector with respect to the base, have necessarily been changed.



P-type point-contact transistor

Comparison of Point-Contact Transistor to Triode Vacuum Tube

The operation of a point-contact transistor can be compared to the operation of a triode vacuum tube. The emitter is equivalent to the grid, the base to the cathode, and the collector to the plate. The base current of the transistor is relatively large, and the major portion of it goes to the collector. Unlike the grid of a vacuum tube, the emitter draws a continuous current.

In the point-contact transistor, the input (emitter) impedance is low and the output (collector) impedance is high. In a vacuum tube the (grid) impedance is high and the output (plate) impedance is low. A comparison of point-contact-type transistors and vacuum tubes is made in the table below:

Vacuum Tube	Point Contact Transistor
Cathode	Base
Grid	Emitter
Plate	Collector
Voltage amp	Current amp
$\mu = \Delta E_p / \Delta E_g$	$\alpha = \Delta I_c / \Delta I_e$
Hight input impedance	High output impedance
I_p	E_c
E_p	I_c
I_g	E_e
E_g	I_e
Constant E supply	Constant I supply
Capacitance	Inductance
Large in size	Small in size
High power consumption	Low power consumption
Fragile to shock	Rugged
Reverse bias	Forward bias

A junction-type transistor has already been compared to a vacuum tube. This comparison is given in table form below.

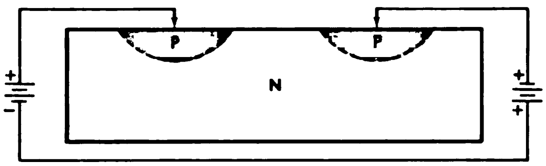
Vacuum Tube	Junction Transistor
Cathode	Emitter
Plate	Collector
Grid	Base
$\mu = \Delta E_p / \Delta E_g$	$A = \Delta I_c / \Delta I_e$
Voltage amp	Current amp
High input impedance (grounded cathode)	High input impedance (grounded emitter)
Low output impedance	Low output impedance
High power consumption	Low power consumption
Fragile	Rugged
Large in size	Small in size

BIAS VOLTAGES

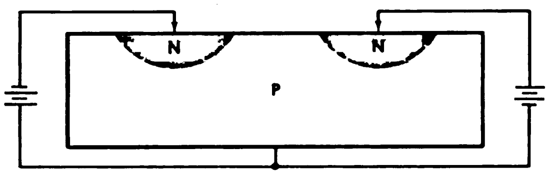
In a vacuum tube the grid is usually biased by means of a voltage which is negative with respect to the cathode. In the case of transsistors, the collector may be positive or negative with respect to the base, and the emitter may be positive or negative with respect to the base. The figures on the next page illustrate the bias polarities required for the various types of transistors.

The bias polarities depend on whether the transistor is an N-P-N or P-N-P junction type, or a "P" or "N" point-contact type.

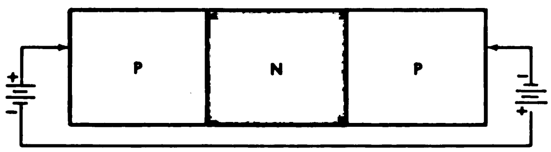
If you remember that the N-type point-contact transistor is, by stretching the imagination a bit, actually a P-N-P type transistor



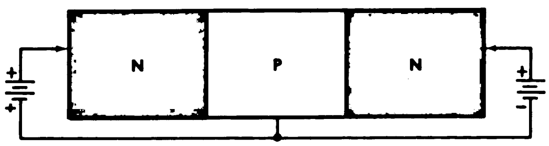
N-type point-contact transistor



P-type point-contact transistor



P-N-P junction transistor



N-P-N junction transistor

and the P-type transistor is an N-P-N transistor, bias polarities can be remembered easily by the following simple rule:

"The emitter is biased in a forward direction with respect to the base, and the collector is biased in the reverse direction with respect to the base, in any transistor."

Although transistors were developed within the past few years, much is expected of them in the immediate future. Many electronic units will be freed from the limitations of the vacuum tube — fragility, bulkiness, short life, and power-supply requirements, which can be overcome with transistors.

Transistors have been successfully adapted to use in such units as phono-amplifiers, radio and television receivers, hearing aids, "walkie-talkies," radar spotting devices, computers, telemetering systems, and other devices which at present employ vacuum tubes.

Vast strides have already been made in overcoming present transistor limitations such as nonuniformity in characteristics, high noise level, and susceptibility to heat and humidity. There is every reason to believe that it is just a matter of time until these difficulties will be surmounted. We can assume that transistors soon will be common components in much of the airborne electronic equipment, especially those devices in which small components are desirable.

Modulation of Carrier Waves

Modulation is a process of impressing intelligence upon a carrier by altering its amplitude, frequency, or phase in accordance with the variations of the speech or signal being transmitted.

The carrier may be a direct current, an alternating current, or a series of uniform pulses repeated at a constant rate.

An unmodulated carrier in itself conveys no intelligence other than that the transmitter is in operation.

When some characteristic of the carrier is caused to vary as a function of the instantaneous value of the modulating signal, the receiver can detect the variation (demodulate the carrier) and translate it into an intelligible form.

If the carrier is in the form of an alternating current or wave, its frequency must be greater than the highest frequency present in the modulating signal in order to include all of the intelligence or variations present in the signal. If the carrier is in the form of a pulse chain, the repetition rate of the pulses must be at least double the highest frequency in the modulating signal.

The frequency of the modulating signal itself is not present in the carrier at any time, but the intelligence produced by the modulation is present in new groups of radio frequencies (sidebands) above and below the unmodulated carrier frequency, or in changes in the time, position, or width of pulses in the pulse chain. These changes in the carrier vary in accordance with the instantaneous values of the modulating signal. Several types of modulation are illustrated on the following page.

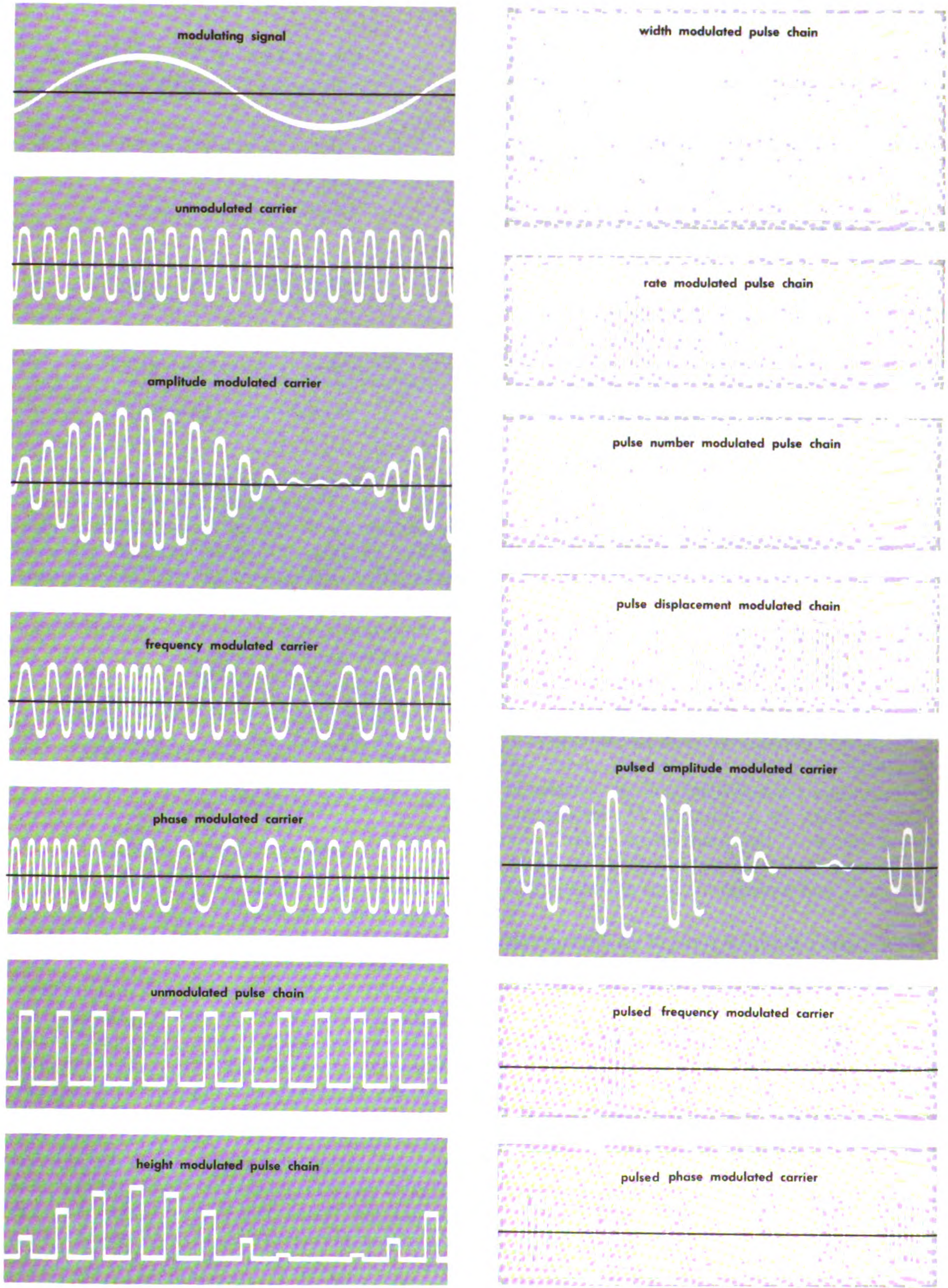
Modulation methods can be classified broadly into five general catagories:

1. Amplitude modulation.
2. Frequency modulation.
3. Phase modulation.
4. Pulse modulation.
5. Doppler principle.

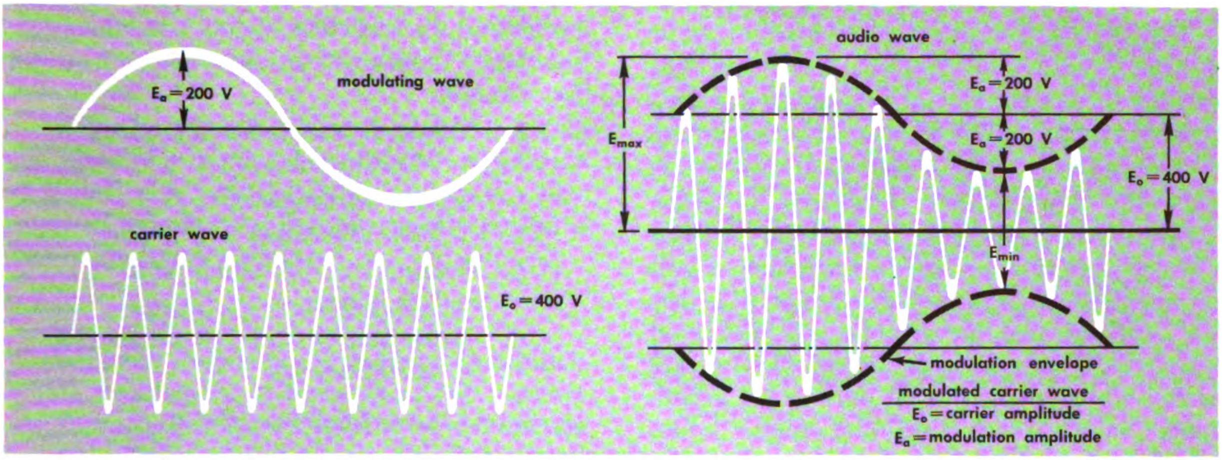
AMPLITUDE MODULATION

Amplitude modulation (AM) is a variation of the radio frequency (RF) power output of a transmitter at an audio rate. This action is illustrated on page 153.

The RF energy of the carrier has to increase or decrease in power in accordance with the audio (signal) frequencies. If the modulating audio frequency is high, the radio frequency (carrier) must vary in amplitude more rapidly than if the audio frequency were low. If the audio signal is loud in volume, the radio-frequency energy must increase and decrease by a larger percentage than if the signal were low in volume. Thus, the radio-frequency-carrier variations must correspond in all respects with the audio-frequency variations to insure fidelity in the transmission of intelligence by amplitude modulation of an RF carrier.



Types of modulation



Graphical representation of amplitude modulation

In analyzing amplitude modulation, it is customary to assume that the modulating signal is sinusoidal in form although it is rarely a pure sine wave.

An RF carrier itself must be free from inherent variations in amplitude such as might be caused by poorly regulated power supplies. It should be as pure a sine wave as possible.

An amplitude-modulated transmitter is designed so that the frequency of the carrier is not affected by the application of the modulating signal, for if this were not done, the carrier might *swing* or shift, which not only would cause interference to other channels but would make its reception by an AM receiver difficult if not impossible.

In the early days of radio broadcasting, this *swinging* of the carrier under modulation was one of the problems which had to be overcome by the transmitter designers and engineers. This undesirable effect in amplitude-modulated carriers later led to the development of two other types of modulation and transmission which we know as *frequency modulation* and *phase modulation*.

The degree of amplitude modulation is expressed by the percentage of maximum amplitude deviation from the normal value of the RF carrier. The effect of an amplitude-modulated wave, as measured by receiver response, is proportional to the percentage of modulation.

The percentage of variation of the total voltage of the final RF amplifier stage in an AM transmitter depends upon the ratio of the audio-frequency (AF) voltage to the DC plate voltage of the RF amplifier. For example, if the DC plate voltage of the RF amplifier is 400 volts and the modulating AF voltage is 200 volts, the two voltages will add (when they are of the same polarity) to give a sum of 600 volts. When they are opposite in polarity, they produce the difference of the two voltages or 200 volts. This example is illustrated in the figure above.

Thus, the plate voltage of the RF amplifier varies between 200 and 600 volts. The transmitter is modulated 50 per cent since the maximum variation (200 volts above to 200 volts below the DC voltage level) is 200/400 or one-half of the value of the DC voltage of 400 volts. Therefore, the percentage or degree of amplitude modulation is defined as the percentage of variation of the modulated wave compared with the unmodulated wave and is expressed by the following formula:

$$m = \frac{E_{max} - E_o}{E_o} \times 100 \text{ or } \frac{E_o - E_{min}}{E_o} \times 100$$

where m = per cent of modulation and

E_o = carrier amplitude (unmodulated)

The amplitude of the carrier should be varied to as high a degree as possible for every variation of the signal in order to produce the highest output possible from the detector in the AM receiver. Varying the amplitude as much as possible is necessary since the output

of the detector varies directly as the amplitude of the carrier (assuming the detector to be linear).

A low-power RF carrier modulated to a high degree can produce a stronger output from the receiver than a more powerful carrier modulated to a lesser degree.

Amplitude modulation must never exceed 100 per cent, because to do this would result in complete cancellation of the carrier at the peaks of the modulating voltage. Cutting off these peaks would produce an interrupted carrier which would cut off part of the intelligence being transmitted and might cause interference to adjacent transmission channels.

The power in the carrier varies either as the square of the current or the square of the voltage, as long as the resistance in the circuit remains constant. This relation in formula is:

$$P = \frac{E^2}{R} \text{ or } P = I^2R$$

Thus, the instantaneous peak power of an amplitude-modulated carrier modulated 100 per cent is four times the power of the unmodulated carrier, because the modulating voltage (or current) is then equal to the carrier voltage (or current), and their sum is equal to twice the carrier voltage (or current). Applying the power formula,

$$P = \frac{E^2}{R} \text{ or } P = I^2R$$

and solving for the value of the power under this condition, you will find the result to be "4P," as follows:

$$\frac{(2E)^2}{R} = 4P \text{ or } (2I)^2 \times R = 4P$$

For maximum power and fidelity of signal transmission, the AM transmitter should be capable of being modulated up to 100 per cent and should have linear characteristics throughout its entire modulation range; that is, the peak power of the carrier at all times should vary as the square of the sum of the carrier voltage (or current) and the modulating voltage (or current).

The amplitude-modulated wave is a combination of DC current, the fundamental RF carrier frequency, the sum and difference frequencies of the carrier, and the modulating signal plus harmonics of each. For example, if the carrier fundamental frequency is 500 kilocycles and the audio frequency is 10 kilocycles, the wave contains the following principal frequencies:

a. Fundamental carrier frequency (f_o): 500 kilocycles
b. Second harmonic of f_o: 1000 kilocycles
c. Sum of f_o and modulation f: 510 kilocycles
d. Difference of f_o and modulation f: 490 kilocycles

Upper and Lower Sidebands of AM

The plate circuit of the final RF amplifier of the AM transmitter, broadly speaking, is tuned to the fundamental RF carrier frequency and transfers the sum and difference frequencies to the antenna. But the circuit discriminates against the lower audio frequencies and the higher-frequency harmonics. Thus, an antenna radiates a wave composed of three frequencies which are close in value. These frequencies are the fundamental carrier frequency and the sum and difference frequencies of both the carrier and the modulating signal. The sum and difference frequencies are called the upper and lower *sidebands*.

The higher the frequency of the modulating audio signal, the farther the sidebands will be from the carrier. If the bandwidth of the transmission is limited, the frequency of the modulating AF necessarily is limited as it is in AM radio broadcasting where the station is limited to a 10-kilocycle channel. A 5-kilocycle AF signal is the highest frequency which can be broadcast by the conventional AM transmitter. A higher frequency would not contain the sidebands within the 10-kilocycle channel. That is, the carrier frequency plus 5 kilocycles and minus 5 kilocycles constitute a separation of 10 kilocylces between the sidebands.

It is possible to broadcast audio frequencies up to nearly 10 kilocycles by special methods such as single-sideband transmission, but certain difficulties are encountered in such systems. These difficulties generally have restricted their application in commercial broadcasting. If the carrier frequency is not raised and lowered simultaneously and inversely with

the modulating frequency, the modulated carrier overlaps into the adjacent channels when the modulating frequency exceeds 5 kilocycles.

It was mentioned earlier that the undesirable effect of swinging that occurs in amplitude modulation led to the development of other types of modulation. Frequency modulation is one of the types developed.

FREQUENCY MODULATION

Frequency modulation is the process through which intelligence is transmitted by varying the frequency of the RF carrier in accordance with the frequency of the modulating signal.

In frequency modulation (FM), the frequency of the RF carrier when it is not being modulated is referred to as the *center frequency, normal frequency,* or *resting frequency* (f_r), the latter designation being most commonly used.

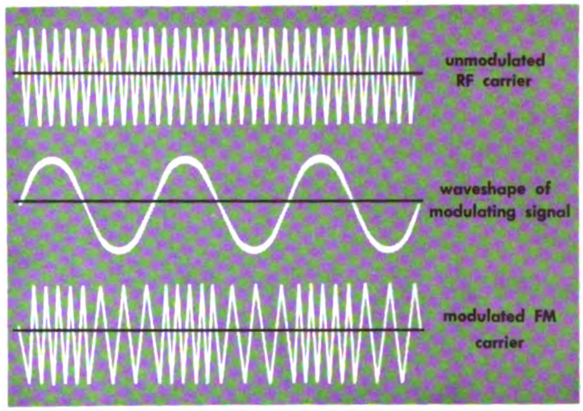
When the carrier is modulated by a positive signal voltage, its frequency increases in proportion to the amplitude of the positive signal voltage. And when the carrier is modulated by a negative signal voltage, its frequency decreases in proportion to the amplitude of the negative signal voltage.

The figure to the right above is a graphical representation of frequency modulation. It illustrates an unmodulated carrier in which each RF cycle occupies the same amount of time, a sinusoidal modulating signal, and a modulated carrier.

The maximum frequency change from the resting frequency is governed by the amplitude of the modulating signal and is called the *deviation.* While the modulating signal does not vary the amplitude of the carrier, it does shift the frequency and concentrate the power in new frequency sidebands.

Rate of Deviation of a Signal

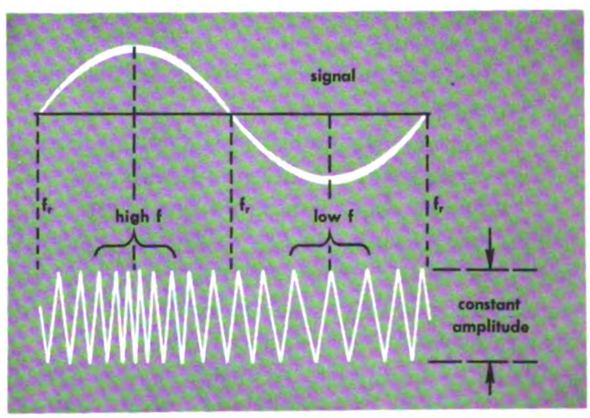
A sinusoidal signal voltage causes an FM transmitter to radiate a signal which changes in frequency from the carrier resting frequency to a higher frequency, back to the resting frequency, then to a lower frequency, and back to the resting frequency in accordance with the frequency of the modulating sine wave. This process, known as the *rate of deviation,* is illustrated on the right.



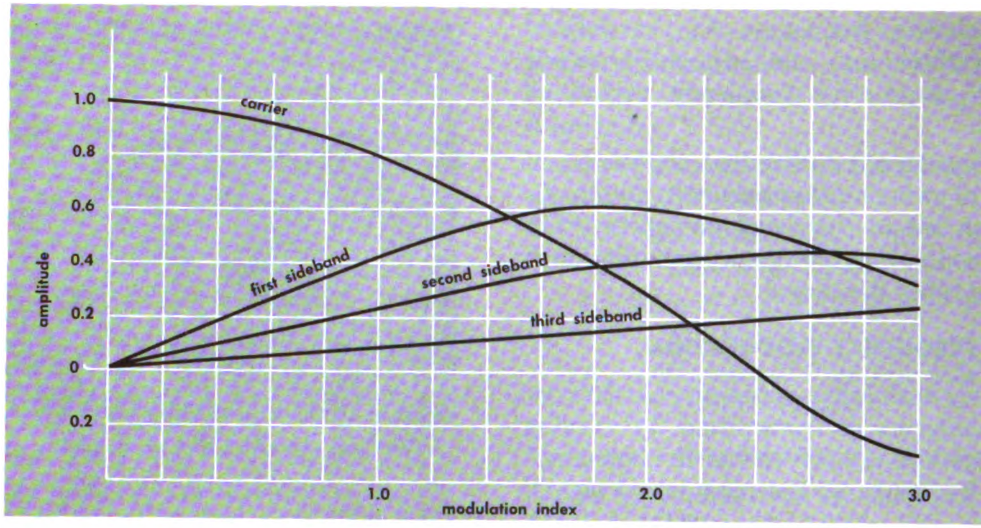
Graphical representation of frequency modulation

An FM transmitter circuit which is designed to produce a deviation of 40 kilocycles can produce signals that differ from the resting frequency by amounts much greater than 40 kilocycles. The circuit can do this because in addition to the normal upper and lower frequency sidebands which are produced by the amplitude of the modulation frequency, other sidebands differing from the resting frequency by multiples of the modulating frequency also are present.

The spread of these frequencies becomes greater with the higher modulation frequencies such as are required in carrier telephone transmissions. This characteristic is inherent in the circuit employed. The strength of the emission on frequencies beyond the normal deviation, in general, does not decrease in a manner directly proportional to their separation from the resting frequency; therefore,



Carrier after frequency modulation



Variation of signal strength with modulation index

the strength of the sidebands at frequencies considerably remote from the resting frequency may be greater than that of frequencies close to the frequency of maximum deviation.

These sideband frequencies can cause interference on other radio channels which may be quite remote in frequency from the interfering transmitter. This interference is most likely to occur with transmitters using very high indexes of modulation.

Modulation Index

The *modulation index* of an FM transmission is comparable to the *percentage of modulation* which is applied to amplitude-modulated signals, even though the index is quite different in character. The modulation index often is defined as the "ratio of deviation of the carrier frequency in relation to the frequency of the modulating signal." It can be expressed by the equation:

$$\beta = \frac{\text{frequency deviation of the RF carrier}}{\text{frequency of the modulating AF}}$$

where β (beta) = modulation index

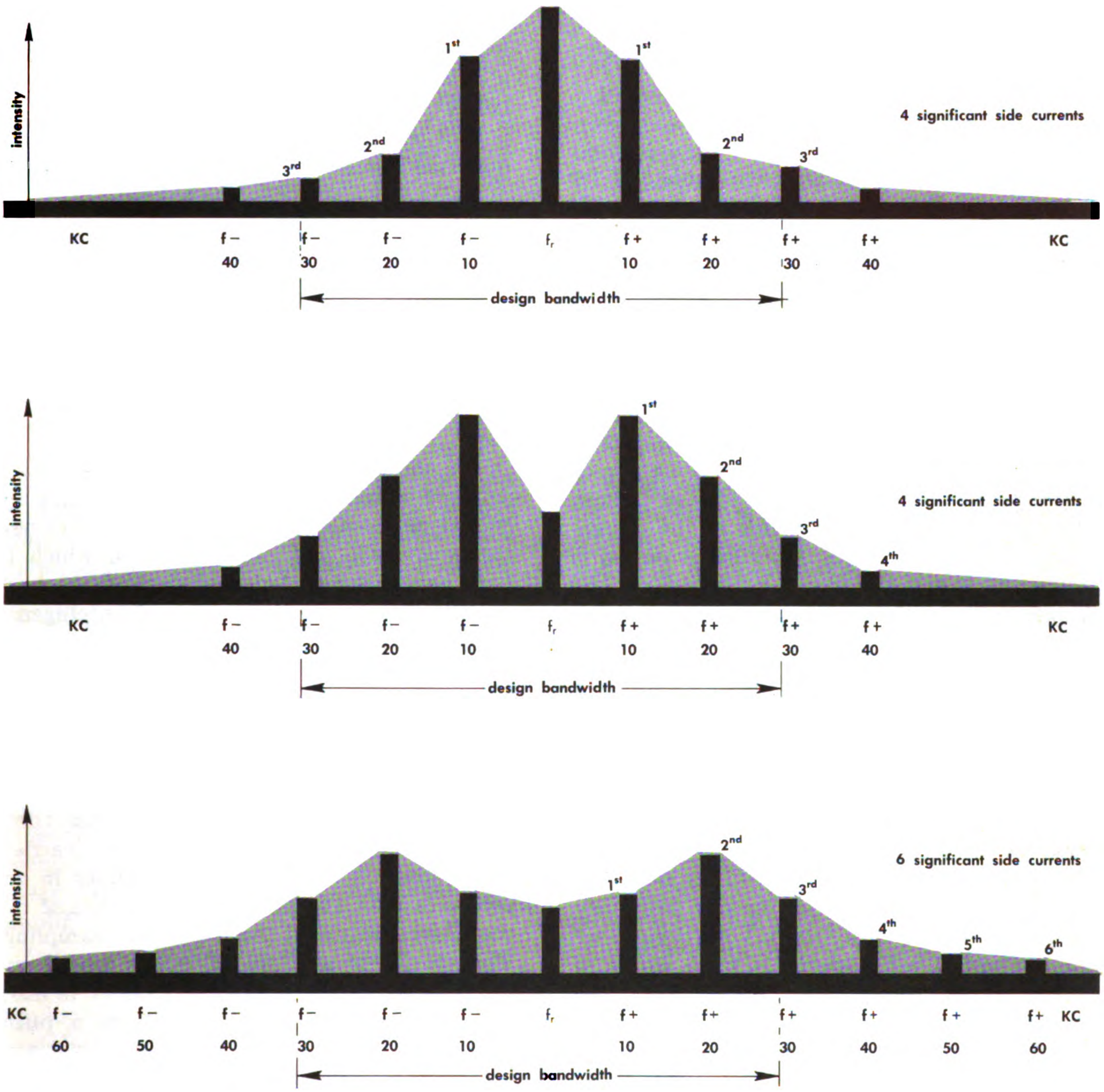
The *limiting modulation index* is called the *deviation ratio,* which is the ratio of the *maximum* carrier-frequency deviation to the *highest* modulating frequency used.

While in amplitude modulation the carrier strength is constant and only the sideband amplitude varies, in frequency modulation the carrier strength varies with the modulation index, as shown in the accompanying graph.

At a modulation index of approximately 2.4, the carrier disappears entirely. It then reverses phase (as compared to its phase without modulation), thus becoming negative. It remains negative until the modulation index is increased to approximately 4.6, at which point it again passes through zero, reverses phase, and becomes positive once more.

If the curves shown in the preceding graph were extended to higher indexes of modulation, additional sidebands would be developed. And the carrier would continue passing through periodic reversals in phase. It would continue crossing zero points at specific indexes of modulation. The mathematical solutions for this behavior are too complex to be treated here, but they can be visualized if you consider a number of sine waves of different frequencies adding and subtracting at certain points to cancel or reverse the phase of any particular wave.

In frequency modulation and phase modulation (PM), which is discussed next, the energy that goes into the sidebands is taken from the carrier. However, the total power in the carrier and its sidebands remains the same regardless of the modulation index. More power may be in the first sideband than in



Spectrum distribution for modulation index

any other sideband at one value of modulation index; the greatest amount of power may be in the second or third sidebands at some other value of modulation index.

Both FM and PM systems possess the advantage of being relatively unaffected by atmospheric or man-made interferences, which are essentially variable in amplitude. But both FM and PM radiations are more or less like light waves in character as a result of the high carrier frequencies employed.

FM is more generally used than PM because it is capable of handling more intelligence (producing more variations in frequency) within a channel of specific frequency limits.

The same receiver can be used for both FM and PM, because the discriminator or *frequency detector* responds to changes in phase in the same manner that it responds to variations in frequency.

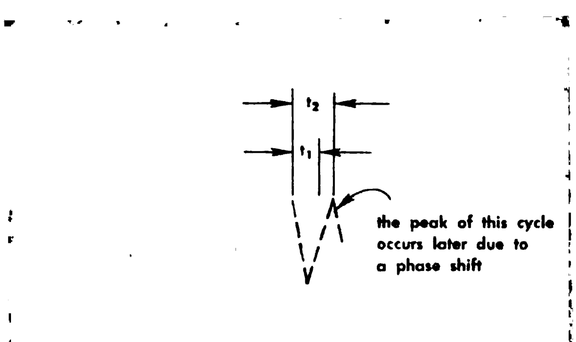
PHASE MODULATION

Phase modulation is an indirect method for obtaining a frequency-modulated carrier. It is accomplished by passing the RF and AF signals through various phase-finding networks which tend to change the phase of the AC radio-frequency carrier voltage.

Thus, the phase of a carrier cycle shifts with respect to each preceeding cycle, and the last cycle is either ahead of or behind its normal phase. In other words, the carrier cycle reaches its peak value either earlier or later than it does in the unmodulated carrier. This phenomenon is illustrated in the accompanying figure in which the normal period is shown as t_1 and the longer period as t_2. The longer period corresponds to a lower frequency since

$$f = \frac{1}{t}$$

Theoretically, the only difference between frequency-modulated waves produced by PM and FM systems is in the phase relationship between the modulating signal and the deviation. But, practically, there are other differences. One difference arises from the fact that with phase modulation, the carrier frequency is much more stable than with frequency modulation. The frequency is more stable in the phase-modulation system because the oscillator can be crystal controlled and the modulating frequency applied to a later stage.



Phase modulation of a carrier

In the FM system, the modulating signal must be applied directly to the oscillator circuit. Since the oscillator is designed to be resonant at the resting frequency, it offers a different impedance to frequencies above or below the resting frequency; therefore, its output varies somewhat in amplitude at frequencies other than the resting frequency.

Another basic difference between FM and PM is that in the FM system the deviation is determined only by the amplitude of the modulating signal, but in the PM system the deviation is proportional to both the frequency and the amplitude of the modulating signal.

PULSE MODULATION

The increasing demands for communication services in both military and commercial applications have led to the development of new types of radio and radar systems, which in turn have necessitated new methods of modulation for the transmission of intelligence in varied and complex forms. *Pulse modulation* with its many possible variations has proven most practical in many of these applications.

Fundamentally, pulse modulation differs from other types of modulation in that the intelligence to be transmitted is sampled during brief, periodic intervals, and these samples are used to modulate the carrier. The carrier is varied in some manner in accordance with the instantaneous value of the modulating signal at the moment of sampling. The carrier is a chain of pulses which are uniform in nature and generated at a fixed rate. The carrier is modulated by a pulse chain from the sampled data which serves as a subcarrier.

Pulse modulation is valuable for applications which require multiplexing or simultaneous transmission of more than one data signal on a common carrier, such as telemetering, wire or micro-wave television cables, oceanic cables, and others. Pulse modulation permits the transmission of many types of data in short periods of time and with a minimum of equipment.

Simultaneous transmission of multiple channels is accomplished by one of two systems. One system involves using a separate *subcar-*

rier frequency for each channel, and the other involves transmitting data samples from each channel in a specific time sequence.

In the first system, known as *frequency division*, each data signal modulates the subcarrier assigned to its specific channel and is identified by the frequency of the subcarrier. In the second system, known as *time-division multiplexing*, the instantaneous amplitude of the signal is sampled from one channel at a time and transmitted in a regular sequence until all channels have been sampled; the process is then repeated in the same sequence until all desired data has been transmitted.

The nature of the data to be transmitted determines the band width required for the transmission and frequency of sampling. For voice transmission, the rate of sampling must be sufficiently rapid to prevent the listener from detecting the intervals between the increments of data transmitted. A sampling rate of approximately 8000 samples per second is sufficient to produce the effect of normal continuous sound to the listener.

Since only one instantaneous value of the modulating signal is transmitted through one channel at any given instant in the time-division multiplexing system, there is no cross-talk or interchannel modulation. This interchannel modulation might occur in the frequency-division system because of the nonlinear frequency response of the modulation amplifiers.

Pulse modulation readily lends itself to multiplexing by time-division because it must employ instantaneous sampling. When pulse modulation is combined with time-division, the system is known as *pulse-time multiplex* and possesses many characteristics which are desirable in a communication system for handling complex data. Some of these characteristics are:

a. High signal-to-noise ratio, which is made possible through the use of limiting and clipping circuits.

b. The "on-off" nature of pulse modulation makes it adaptable to simple repeater systems for increasing range of transmission.

c. Eliminates the need for complex filter networks.

d. Adaptability and flexibility of application.

e. Freedom from interference due to interchannel cross-modulation.

Pulses have individual characteristics (parameters) with respect to height, width, duration, repetition rate, formation time, decay time, shape, and displacement from normal occurrence. These characteristics may be utilized singly or in combinations to provide a wide variety of pulse-modulation methods adaptable to the transmission of intelligence in many forms and degrees of complexity.

In pure pulse modulation, both the rate and timing of the signal sampling remain constant regardless of how the pulses in the chain are modulated.

When the pulse-modulated system is used for transmitting intelligence from several sources, or data of a highly variable nature, the ratio of noise to signal may become excessive (as is true in systems using conventional types of modulation). Signal-to-noise ratio may be improved through the use of clipping and limiting circuits and by increasing the band width, if possible, to permit wider separation of individual data channels. The transmission band width depends upon pulse shape as well as pulse rate, and the method and degree of modulation.

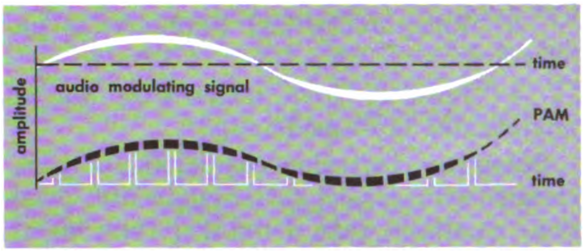
The fact that in pulse modulation energy is transmitted only for short intervals makes possible the use of high peak power in the carrier and the use of magnetrons and klystrons to generate carriers of very high frequency.

When high-frequency carriers are used, the band width can be made much greater and more data channels can be accommodated. It is customary to interrupt the carrier between modulation pulses, but this is unnecessary except to affect a slight saving in power consumption.

We now will consider several types of pulse modulation, taking up pulse-amplitude modulation first.

Pulse-Amplitude Modulation

When the amplitude of a pulse is modulated, the transmission is referred to as *pulse-ampli-*



Pulse amplitude modulation

tude modulation (PAM). Although easily applied, this system is not widely used because it requires equipment with linear characteristics. Too, it can operate at an average of only 50 percent of its maximum capabilities. The figure above illustrates a pulse-amplitude modulated pulse chain.

In PAM systems, either of two types of amplitude modulation are used. One of these is known as *unidirectional* PAM. This type employs pulses of one polarity; the other type is known as bidirectional PAM because it employs pulses which vary in amplitude above and below a fixed reference level. The pulses are positive and negative in amplitude with an average value equal to zero. The illustration below shows two types of pulse-amplitude modulation.

Pulse-Width Modulation

Pulse-width modulation (PWM) occurs when the width or duration of the pulse varies symmetrically in accordance with the modulating signal or when the leading or trailing edges of the pulses are modulated. Pulse-width modulation is commonly used on modulating and demodulating other pulse modulations. If PWM is differentiated or *gated*, *pulse-displacement modulation* is obtained. If a width-modulated pulse chain is passed through a low-pass filter, the signal is removed directly.

Pulse-width modulation is a convenient medium for transmitting intelligence but involves a variation of duty cycle that reduces the operating efficiency of the equipment. A combination of pulse-width modulation and pulse-rate modulation could be utilized to hold the duty cycle constant. The cycle is held constant by automatically decreasing the rate of wide pulses and increasing the rate of narrow pulses at rates inversely proportional to the width of the pulses.

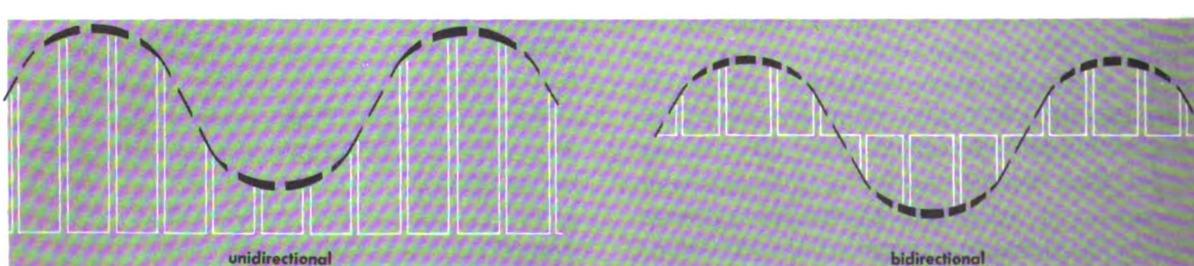
The *duty cycle* is the ratio of the pulse length (duration) to the pulse-repetition period. To convert peak power into average power, peak power is multipled by the duty cycle. These terms are illustrated in the figure at the top of the next page.

In pulse-width modulation, the variation of average power in the carrier is approximately the same as in pulse-amplitude modulation, but the amplitude of the pulses remains constant. The constant amplitude of the pulses facilitates the limiting and clipping of the pulses to eliminate extraneous noise, as shown in the series of figures to the right.

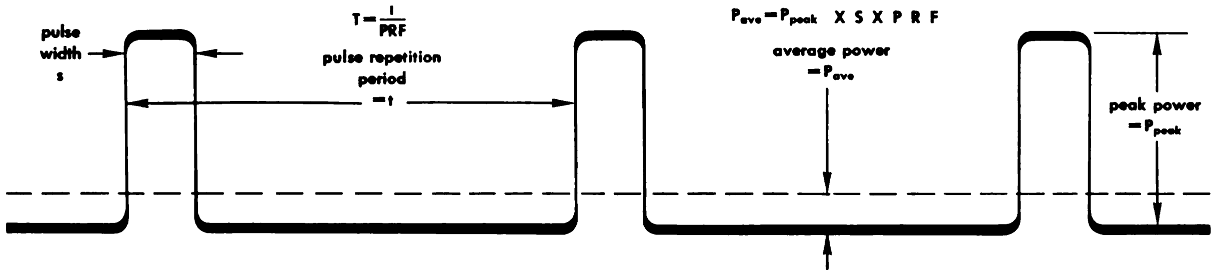
Pulse rate modulation also is subject to variations of duty cycle and loss of average power during the modulating cycle. It is not commonly used because of this drawback.

Pulse-Displacement Modulation

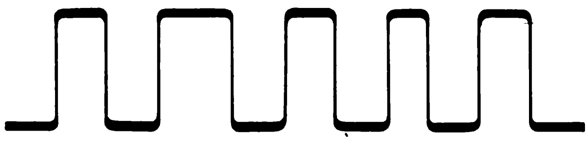
Pulse-displacement modulation, also called *pulse-time* and *pulse-position* modulation, is widely used in microwave time-division multiplex because it permits interlacing several pulse chains without confusion. Other pulse modulations offer similar possibilities if the extent of modulation is limited.



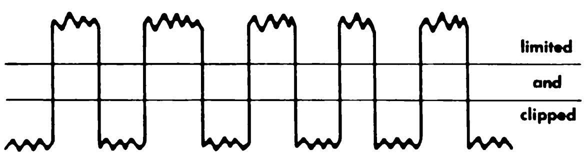
Types of pulse amplitude modulation



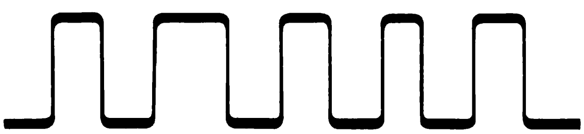
Definition of pulse characteristics



Original train of width-modulated pulses



Pulses with addition of noise



Amplified output of limiter and clipper

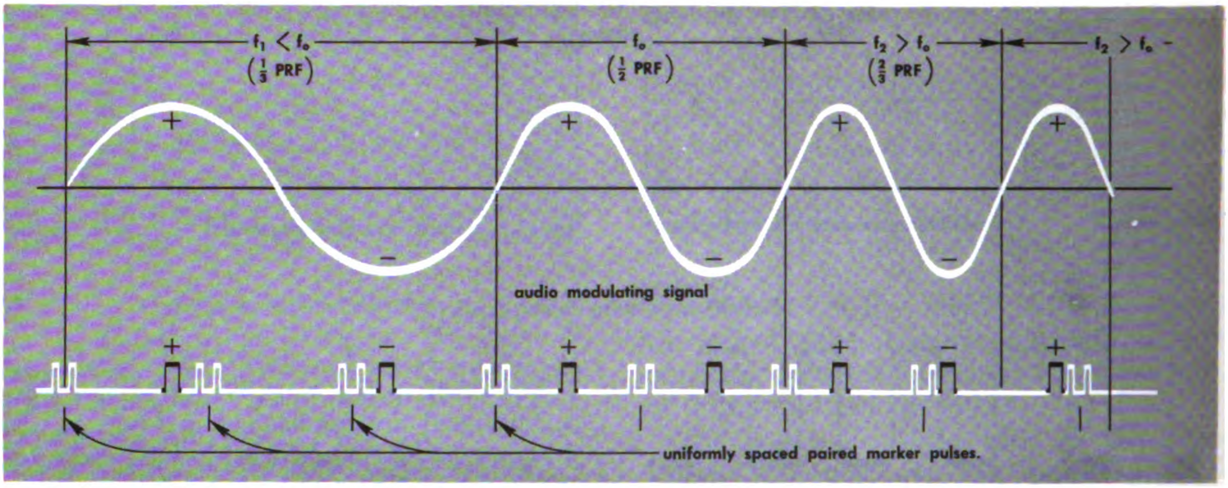
Pulse displacement modulation (PDM) is accomplished by varying the time between pulses or by varying the displacement of the signal pulse with reference to a *marker pulse.* A marker pulse is supplied from a separate marker generator, such as a free-running multivibrator. The marker generator modulates the carrier at uniform intervals with pulses of constant width and amplitude, or with pairs of pulses which are readily distinguished from those produced by the signal.

There are several methods by which the audio frequency signal can be applied to produce pulse displacement modulation. One method employs a driven blocking-oscillator. The blocking-oscillator conducts when the positive-going portion of the audio signal voltage is applied to its control grid or when the negative-going portion of the audio signal is applied to its cathode.

The positive and negative portions of the audio signal are applied successively to the grid and cathode of a single blocking oscillator through limiting or clamping circuits; or two properly biased blocking oscillators are used, one of which is biased to conduct only on the positive portion of the audio signal, while the other is biased to conduct only during the nagative portion. In either case, the end result is the same. The blocking oscillator conducts and produces a modulating pulse when the amplitude of the input signal attains a predetermined positive and/or negative value. The frequency of the audio signal in this instance, determines the frequency with which the modulating pulses are generated. When these pulses are superimposed upon, or used to trigger, a chain of carrier pulses which is being modulated at uniform intervals by a *marker pulse* (or pair of pulses), the position of the signal pulses with respect to the marker pulses varies in accordance with the frequency variations of the signal. Thus, the intelligence conveyed by the signal is represented in terms of the relative position between the signal pulses and the marker pulses, as illustrated in the top figure on the next page. In the figure, the audio modulating signal is represented by three sinusoidal waves, each of a different frequency.

For illustrative purposes, the frequency of the middle sine wave is represented as being equal to one-half the pulse recurrence frequency (PRF) of the *paired marker pulses.* Thus, each half of the audio cycle produces a pulse which is exactly midway between the adjacent paired marker pulses. The two signal pulses are separated by the same distance as the marker pulses.

The first sine wave represents a signal of lower frequency (one-third of the marker PRF



Example of pulse displacement modulation

in this example), and the distance between the signal pulses is correspondingly greater. No signal pulse appears between one pair of marker pulses. The third and fourth sine waves represent a signal of higher frequency (two-thirds of the marker PRF as shown). The distance between the signal pulses is correspondingly less, with two signal pulses appearing between one alternate pair of markers and one signal pulse appearing between the other alternate pair of markers. In other words, within two *marker intervals* three signal pulses appear.

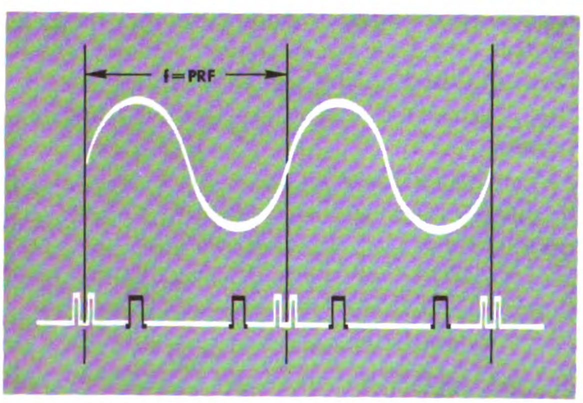
If the audio signal were of the same frequency as the marker PRF, two signal pulses, one for each half of the audio cycle, would appear between each successive pair of markers, as shown below.

Pulse displacement modulation can also be limited and clipped to reduce noise, but it gives less average carrier power output than pulse width modulation for the same peak power.

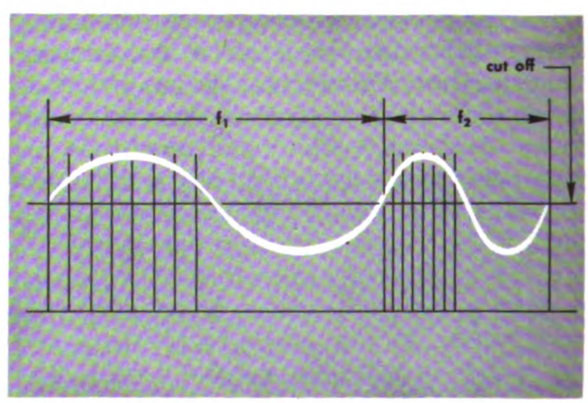
Pulse Frequency Modulation

Pulse frequency modulation (PFM) is a system in which the frequency of the carrier pulses is varied in accordance with variations in the amplitude or frequency or both of the modulating signal. The average power in the carrier for both pulse displacement modulation and pulse frequency modulation is fairly constant.

Pulse frequency modulation is quite similar to pulse displacement modulation except that no separate marker pulses are employed, and



Pulse displacement modulation produced when modulating signal frequency is equal to PRF of paired marker release



Pulse frequency modulation produced by audio signals f_1 and f_2, with f_2 equal to 2f

the variations of the audio signal are used to produce corresponding variations in the PRF of the carrier pulses. The development of *frequency-modulated magnetrons* has made the pulse frequency modulation system practical for many applications which require the transmission of complex data in which the signal may vary greatly in frequency or amplitude within short intervals of time.

The principle of pulse frequency modulation is illustrated in the figure to the left below. Only the positive portion of the audio signal is used for producing the modulation which is in the form of groups of pulses. The frequency of the pulses contained in a group varies in accordance with the frequency of the audio signal.

In the example, only the positive half of each audio cycle is used to produce modulated pulse groups. This procedure is followed so that the groups can be separated by unmodulated intervals for ease of interpretation and decoding.

The above type of pulse frequency modulation is accomplished by applying the audio signal to a multivibrator type modulator or to a magnetron oscillator circuit. In the case of the latter, an FM-CW type magnetron modulation is employed. Modulation is accomplished in such a manner that the field strength of the magnetron (and consequently its frequency) is varied in accordance with the variation in frequency of the audio signal.

Multivibrator type modulators are now most commonly used in pulse frequency modulation systems but are being supplanted by the FM-CW magnetrons wherever practicable, because the latter are more reliable in operation and give an output with a much higher signal to noise ratio.

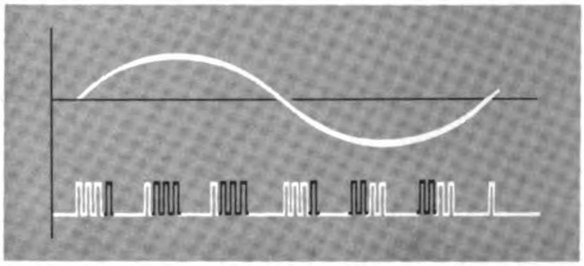
Pulse Count Modulation

It is possible to use the variations in the modulating signal to produce groups of pulses varying in number and concentration with respect to a *normal group count* without modulation. This process is called *pulse count modulation* (PCM).

Pulse count modulation is produced by utilizing the signal voltage to vary the bias of a *keying circuit* which in its normal (no signal) unmodulated operation produces groups of pulses. The pulses are uniform in number and concentration. The keying circuit may contain a "one-shot" multivibrator.

The keying circuit is designed so that a positive-going signal voltage will increase the numer and/or concentration of the pulses in the normal group. Also, a negative-going signal voltage will decrease the number and/or concentration of the pulses in proportion to the variations in the amplitude or frequency, or both, of the audio signal.

The principle of PCM is illustrated in the following figure.



Pulse count modulation

Pulse count modulation is quite similar to pulse frequency modulation except that in the PCM system, the frequency of the carrier is not altered. Too, in the PCM system, intelligence is conveyed by the number of pulses contained in each *signal-modulated group* as compared to the number of pulses in the *normal group*.

Pulse Modulation Thus Far

Some types of pulse modulation, such as pulse amplitude modulation, can be detected by conventional circuits employing crystal-diode mixers and detectors. But more complex types, such as pulse displacement, pulse width, and pulse count modulation, require special synchronizing and decoding circuits. These circuits contain delay and coincidence stages, integrators, pulse-width, and PRF discriminators in addition to the crystal-diode mixers and detectors. The circuitry and principles of these special units are covered in detail later under Coding and Decoding Systems.

	Amplitude Modulation	Frequency Modulation	Pulse-Displacement Modulation
SIGNAL LEVEL FROM TRANSMITTER	Varies with level of modulation.	Remains constant during modulation.	Remains constant during pulses.
AMPLITUDE OF MODULATING VOLTAGE	Determines instantaneous change in signal level. The stronger the signal, the greater the instantaneous change in carrier level.	Determines instantaneous deviation in frequency from carrier resting frequency. The stronger the audio signal, the greater the frequency deviation of the carrier.	Determines instantaneous deviation in time of channel pulse from rest position. The stronger the audio signal, the greater the time deviation.
MODULATING VOLTAGE FREQUENCY	Determines rate of change of amplitude of RF wave.	Determines rate at which carrier frequency changes between high and low values.	Determines number of samples transmitted for each cycle of modulating voltage.
SIDE BANDS TRANSMITTED	Width of transmitted side bands determined by frequency of modulating voltage. Present general limit — 5 kc each side of carrier.	Width of transmitted side bands determined by the amplitude of the modulating voltage. Present limits in military are 40 kc each side of rest frequency. In addition, a 20-kc guard band is provided for separation of adjacent channels. Divided equally on both sides of resting frequency.	Width of transmitted band is determined by pulse width of transmitted pulses. Bandwidth is several times that required by AM or FM. The band width to transmit the same amount of intelligence may be approximated by formula: Bandwidth in mc = 1 pulse length in microseconds.
MODULATOR POWER	One-half plate power input to modulated stage.	Negligible enough to supply plate power loss in modulator tube.	Negligible enough to supply plate power loss in modulator tube.
CARRIER POWER	Final amplifier must be capable of supplying four times the rated carrier power on 100 per cent modulation peaks.	Final amplifier must be able to supply rated carrier power only.	Final amplifier must be able to supply rated carrier power only.
FREQUENCY LIMITATION	None.	Normal above 20 mc, practicable at 2 mc.	Above 1000 mc.
MODULATION SAMPLING	Continuous.	Continuous.	Periodic.

Comparison of modulation methods

Crystal diodes have supplanted vacuum tubes as mixers and detectors in many UHF circuits because of their negligible transit time and better signal-to-noise output ratio.

In the conventional diode vacuum tube, the physical space between the cathode and anode is large compared to the wavelength of the UHF carrier, and the transit time required for electrons to pass from cathode to anode limits the frequencies at which the tube may function as a detector. Crystal diodes are relatively free from these limitations, and thermal noises generated within the crystal are negligible. Hence a much higher signal-to-noise ratio may be obtained from the crystal, making it much more desirable for the demodulation of complex signals.

In many radar systems, 100 per cent pulse amplitude modulation is used because it is the simplest method for modulating magnetron and klystron type oscillators, and it produces the highest peak power in the carrier. Also, at ultra high frequencies, the intensity of static interference becomes less appreciable, making the advantages of the frequency-modulated and phase-modulated systems less apparent.

Pulsed phase modulation is another method for achieving the same result as obtained by pulse frequency modulation, PPM is more suitable than PFM with equipment in which the signal pulses are applied directly to the carrier oscillator.

Although bandwidths are wide at ultrahigh frequencies and selectivity normally would be high, the frequency drifts inherent in UHF equipment restrict selectivity and consequently limit the available bandwidth.

In whatever form pulse modulation may be used, as dictated by the nature of the data to be transmitted, it provides the means for obtaining the fullest utilization of the bandwidths available with a minimum of interference from extraneous noises and intercarrier cross-modulation. For these reasons, pulse modulation is becoming more widely used in both military and commercial communication systems.

Before taking up the Doppler principle, look over the table at the left. It gives a compact review of three types of modulation.

DOPPLER PRINCIPLE

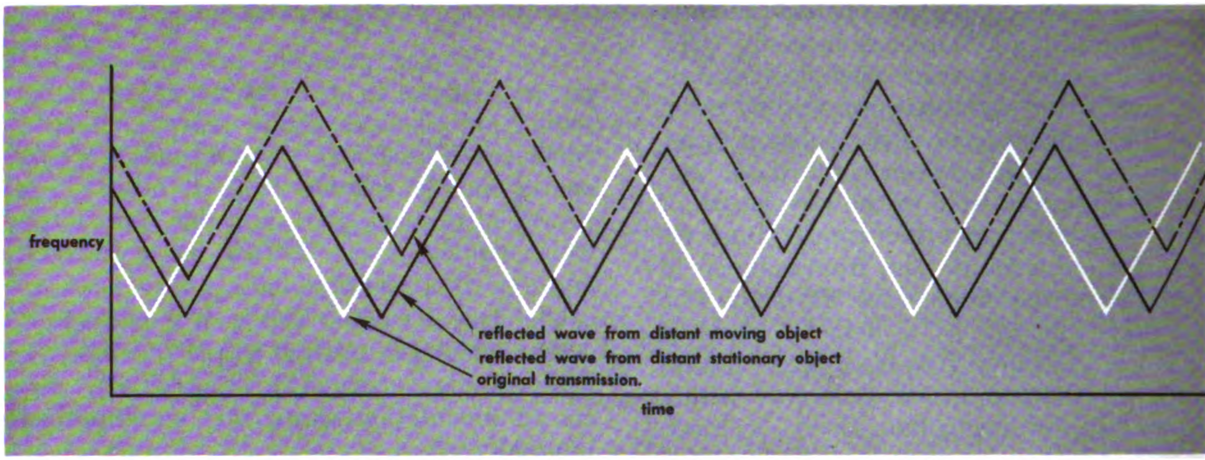
The Doppler principle involves the apparent change in the frequency of light, sound, or electromagnetic waves observed while the source and the observation point are in motion relative to each other. This principle is employed in radar equipment which supplies precise altitude, space-position, and velocity information.

The Doppler effect produces a frequency-shift modulation external to the transmitting equipment as contrasted to previously discussed modulation which was accomplished internally.

In 1842, Christian Johann Doppler of Prague stated that if the distance is changing between an observer and a source of constant vibrations (such as sound or light) the wave number appears to become greater or less than the true value, according to whether the distance is being diminished or lengthened. This effect is commonly observed in the change of pitch of a train whistle as the train approaches and then continues on by.

Electromagnetic waves possess many of the characteristics of light, particularly as their frequencies approach that of light. The Doppler principle is applied to electromagnetic radiations in connection with radio and radar equipment for determining velocity and distance of moving targets such as guided missiles.

In one system which can be used to track a missile, the transmitter operates at a constant and accurately known frequency (f). The beam from the transmitter triggers a beacon transmitter in the aircraft. The beacon transmitter in turn operates at twice the original frequency (2f). The latter signal is picked up by ground receiving stations. At the same time, the ground stations receive the ground transmitter's signal and double its frequency internally through a frequency-multiplying stage. The two doubled frequencies are then mixed. The resulting difference frequency is proportional to the velocity of the missile. The difference is proportional to the velocity because the motion of the missile effectively changes the transmitted frequency by an amount proportional to the velocity of the missile. The velocity vector between the missile and any point on the ground can be



Doppler effect on frequency modulation (sawtooth wave)

accurately measured by means of a *doppler radio* system.

By integrating the velocity values from the instant of takeoff, position vectors can be determined with great accuracy. If two or more stations are used, the position of the missile with respect to a fixed coordinate system can be readily established.

If instead of transmitting a continuous wave, the doppler system transmits a sawtooth frequency-modulated wave, both the velocity and the instantaneous distance between the missile and the ground station can be determined simultaneously.

This principle is illustrated in the graph above. Note that the received wave from a distant moving object is shifted both to the right and upward with respect to the original transmission. Consequently, the beat frequency will be alternately very small and very large on succeeding half cycles. The sum of the two different values of beat frequency thus produced is a measure of the distance of the missile from the ground station, and the difference between the two values of beat frequency is a measure of velocity of the missile with respect to the ground station.

Thus frequency-modulated radar determines the distance to a reflecting surface by measuring the frequency shift between transmitted and reflected waves.

While the wave is traveling to the surface and back, the transmitter frequency is changing under the influence of frequency modulation. When the reflected wave arrives back at the transmitter, its frequency is slightly different from the frequency being transmitted at that instant. The transmitted and reflected signals are combined in a mixer-detector circuit. The frequency difference between them is developed as a *beat note.*

The frequency difference becomes greater as the distance between the transmitter and the reflecting object increases. Thus, altitude and range distances may be obtained by interpretation of the *beat frequency* between the transmitted signal and the received signal at any given instant of time. Velocity data may likewise be obtained.

This FM Doppler system is capable of measuring short distances accurately, while a pulse radar system is limited by the width of the pulse. With a pulse width of only 0.2 microsecond duration, the minimum range is about 100 feet. Therefore, the FM Doppler system is more practical for such applications as aircraft altimeters and telemetering where relatively short distances and velocities must be measured accurately. The pulsed radar system is suitable for long-range or high-altitude measurements for which purposes it is commonly used.

To employ the FM Doppler system for long-range measurements would require lowering the rate of deviation of the frequency-modulation. But lowering the rate would in turn reduce its accuracy for shorter range measurements.

Time and Frequency Relationships in Modulation Envelopes

The graph below illustrates the time and frequency relationships existing when a *triangular* modulation envelope is utilized in an FM radar system. The transmitted and received signals combine to produce beat notes at audio frequencies.

The solid sawtooth wave represents the transmitted signal, and the dashed sawtooth wave represents the received signal, both plotted as functions of time.

The total deviation of the frequency modulation (the peak to peak value of the modulation envelope) is indicated as ΔC megacycles. If the center-frequency is 220 megacycles and the deviation is plus or minus 400 kilocycles ($\Delta C = 0.8$ megacycles), the carrier frequency then deviates from 219.6 to 220.4 megacycles.

The received signal is frequency modulated by the same envelope as the transmitted signal because the signal preserves its form during reflection, but the received signal is displaced in time by the reflected interval. The reflected interval is equal to 2 d/c, where "d" is the distance in feet from the transmitter to the reflecting surface and "c" is the velocity of radio wave propagation in feet per second (about 984,000,000 ft/sec or 186,000 miles/sec).

As the time displacement occurs, a corresponding frequency displacement results. This is indicated by the vertical separation (Δf) between the solid lines and the dashed lines on the graph.

The relationship between the distance (d) and the frequency difference (Δf) can be determined by comparing similar triangles in the figure. The height of the triangular waveform is ΔC and half its base is $1/(2fm)$, where "fm" is the frequency of the modulation envelope. The ratio of the height to half the base is $2\Delta Cfm$, and this is equal to the frequency difference divided by the time difference:

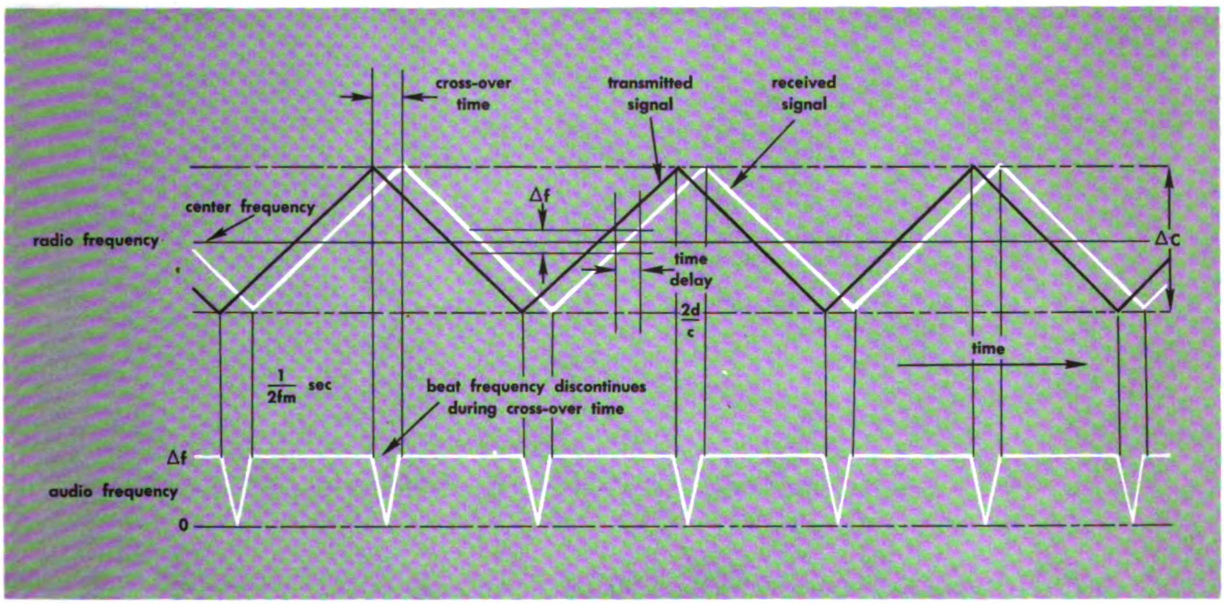
$$2\Delta Cfm = \frac{\Delta f}{2d/c}$$

where ΔC, fm, and Δf are all measured in cycles per second. The relationship between frequency difference and distance is then found to be:

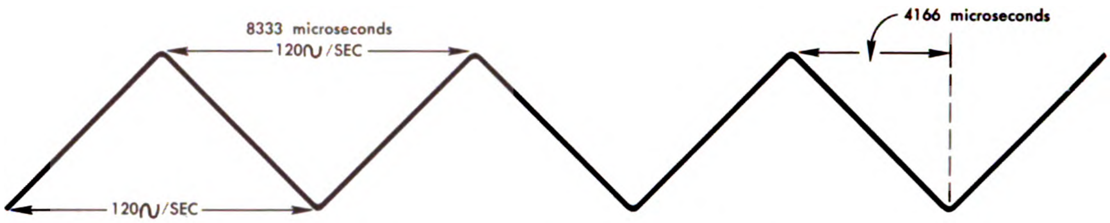
$$\frac{\Delta f}{d} = \frac{4\Delta Cfm}{c}$$

Note that the sensitivity of the indication, in cycles per second frequency-difference per foot, can be increased by using a wide frequency deviation or a high modulation frequency, or both.

If ΔC is 0.8 megacycles, "fm" is 120 cycles per second, and "c" is 984,000,000 feet per



Graphical representation of time and frequency relationships



Time relationship of a 120 cps triangular wave

second, the ratio of frequency difference to distance $\Delta f/d$ is approximately 0.39 cycles per second *per foot of distance*. The equation reads:

$$\frac{4 \times 0.8 \times 10^6 \times 120}{984 \times 10^6} = 0.39$$

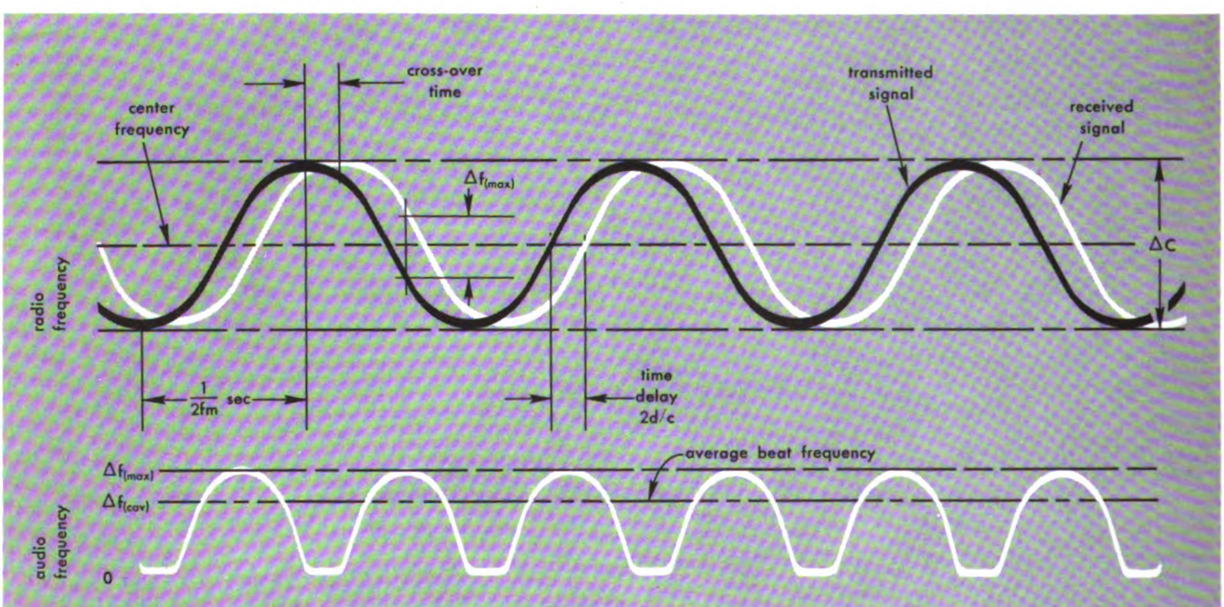
Thus, the maximum beat frequency developed from a reflecting surface 5280 ft. distant is 5280×0.39 or 2059 cycles per second.

A small error occurs during the period between reversal of the transmitted frequency deviation and the corresponding reversal in the received wave. This period is shown in the graph as crossover time. The transmitted frequency, at the center of this period, falls to a value equal to the received signal, which is still increasing at this point, resulting in a beat note of zero at this instant. However, the duration of the *crossover* interval is only a few microseconds per mile as compared with 4166 microseconds, the half-period of the modulation envelope at 120 cycles per second; thus, the error of crossover effect is negligible. These time relationships are shown in the figure above.

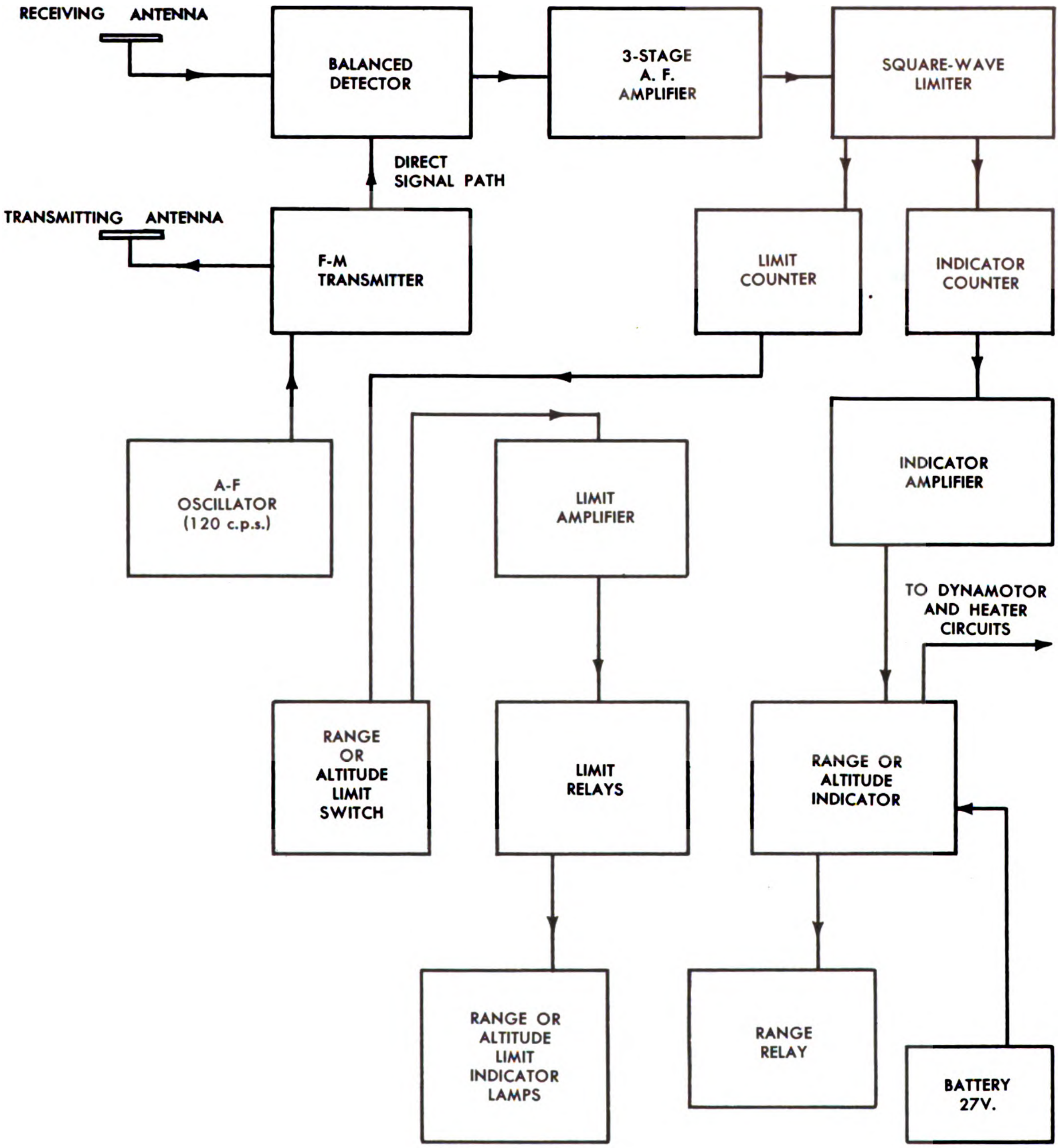
If the modulation is in the form of a *sinusoidal* envelope as in the figure below, instead of a triangular envelope, the frequency difference between cross-overs is no longer a constant. Instead, the difference varies from zero at the cross-over to a maximum which occurs when the transmitted frequency is passing through its center value. By means of suitable circuits, the average beat frequency is measured. This value corresponds with that produced by triangular-wave modulation.

The equation

$$\frac{\Delta f}{d} = \frac{4 \Delta C f m}{c}$$



Beat frequency produced by cross over



Typical air borne FM radar altimeter

is used to predict the average beat frequency (Δf) when sinusoidal modulation is used.

Employing Doppler Principle in FM Radar Altimeter

The functional block diagram above shows a typical airborne FM radar altimeter using the Doppler principle for determining altitude of aircraft.

A typical double-diode counter circuit, which can be used in the indicator counter listed in the preceding block diagram, develops a positive DC voltage proportional to the beat note frequency. This circuit shown on the next page includes a double-diode (12H6), one section of which passes the positive half-cycles of the limiter output, charging the 0.125 uf capacitor and load resistor in shunt. The values of

capacitance and resistance are so chosen that the charge on the capacitor leaks off through the resistor at a rate approximately equal to the conduction of charge through the diode.

Thus, when the positive rectangular waves from the limiter arrive at an increasing rate (higher beat-note frequency), the direct voltage across the capacitor tends to increase. The voltage decreases when the waves arrive at a decreasing rate.

This voltage is passed through a low-pass RC filter which averages the direct voltage and applies it to the grid of the output amplifier tube (12SH7).

A milliammeter in the cathode circuit of the output amplifier tube registers, over a range of 5 milliamperes (1.5-6.5 ma), the average value of the direct voltage on the grid. This meter is calibrated directly in feet. The HI-LO scale shown is switched synchronously with the switch which controls the total deviation of the transmitted signal so that the meter scale corresponding to wide or narrow deviation range is always correctly selected.

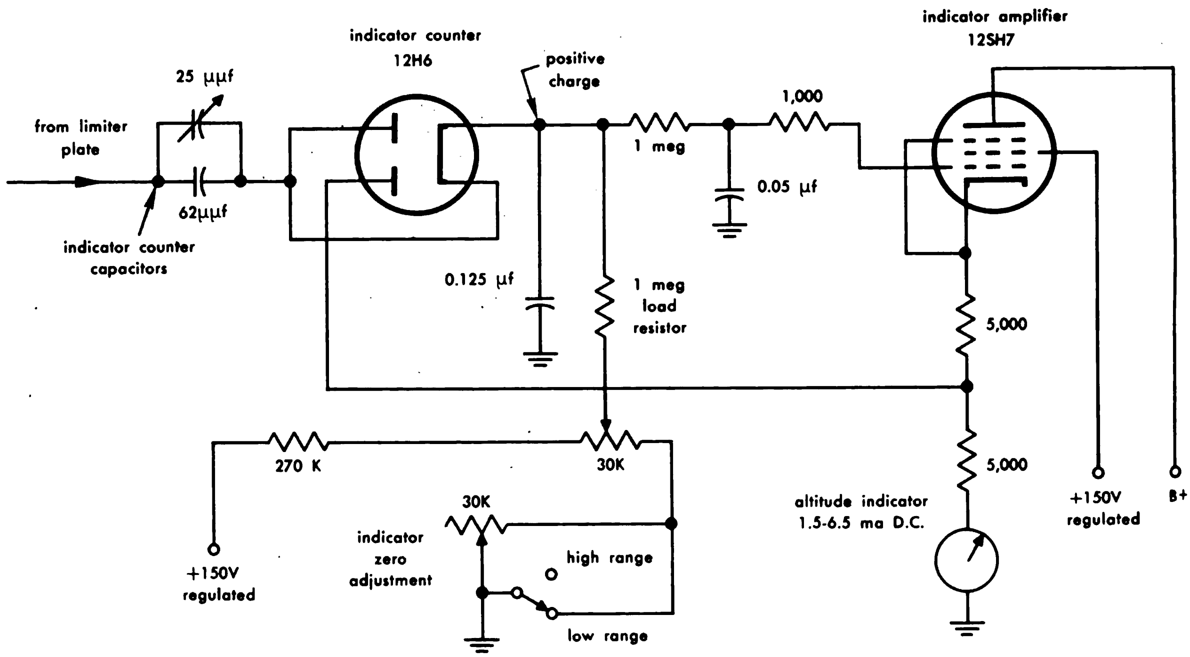
Part of the voltage across the cathode resistance is fed back to the other section of the dual-diode (12H6) in the counter circuit. This diode passes the negative halves of the limiter output to ground. This discharges the coupling capacitor, leaving it ready to receive the next positive half of the limiter output wave. The feedback from the indicator amplifier also biases the diode and makes the indication more linear.

You can now recognize the importance of the Doppler principle to the missile field. As this section has brought out, the Doppler effect is widely used in electronic equipment.

A REASON FOR UNDERSTANDING MODULATION

Since a carrier must be altered in amplitude, frequency, or phase before it can transmit intelligence, such altering becomes vital to the effective use of electronics equipment.

Thus, you should be fully acquainted with the types and functions of modulation; you should understand the phenomenon of modulation before going into the next chapters which take up the detailed study of the components of a missile system.



Double diode counter circuit

Chapter 5

Components of Guided Missile Control Systems

Devices which make up the various types of guided missile control systems are given individual treatment in this chapter. To help you visualize the operation of these components in control systems, a brief mention of overall system operation is covered here. A later chapter goes into more detail on system operation.

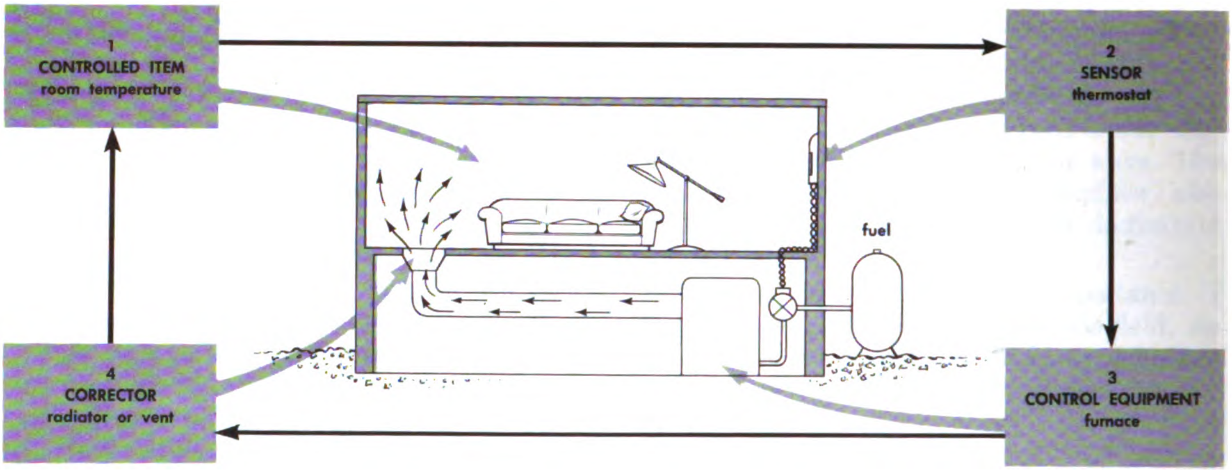
Guided missile control system normally refers to a system that automatically controls the flight of missiles. This system is similar to the automatic pilot systems installed in some large piloted aircraft. The system moves control devices at the proper time and through the distance required to keep the missiles from pitching, changing heading, weaving back and forth, tailspinning, or rolling.

The principles of the missile control system are the same as any automatic control system. *Such a system constantly makes corrections of some controllable item, and then checks the results as a basis for further corrections.* This forms a cycle of interdependent actions which is called a *closed-cycle* control system. Any block diagram of the actions are connected to form what is called a *closed loop,* as shown in later illustrations.

There are four main requirements of any automatic control system. The first requirement of a control system is the presence of *something which is controllable.* For example, it would not be feasible to attempt to control the temperature of the outside atmosphere.

The second important requirement is a means of detecting or *sensing* a departure of the controllable item from the desired condition or reference. This departure is usually called *error.*

A third function is to convert the error information into a form which can be used to regulate the controlling device. The controlling or correcting device is what actually affects the correction of the controlled item



Four parts of every automatic control system

and is the fourth requirement of any system.

One of the most common automatic control systems is used to maintain a certain temperature in a room. A heating system is linked with the four basic system requirements in the accompanying diagram.

The thermostat is adjusted for a certain room temperature. This adjustment represents the *only order* to the system. The temperature for which the thermostat is set then becomes a reference. If the temperature deviates from this reference, the thermostat detects it and sends an error signal to the fuel valve. Heat is produced by the equipment which follows the thermostat. The heat correction is imparted to the room by means of a vent or radiator. Meanwhile, the thermostat, still measuring temperature, insures that the fuel supply is cut off at the proper time to maintain reference temperature.

The figure at upper right shows the same four requirements of an automatic system, applied this time to a guided missile attitude control system. The missile attitude is a controllable item. Deviations may be sensed by a gyro. The controlling equipment changes this signal to a force that can move the control surface the proper amount. The device which actually imparts correction is the control surface (or movable jets, etc.).

The series of events has not stopped. When the missile moves as a result of corrective action, this motion is fed back as information to the sensor, since it is continually detecting attitude. This feedback information completes the *major loop* of the system. This major loop is sometimes called the *dynamic loop* since it includes the motion of the guided missile.

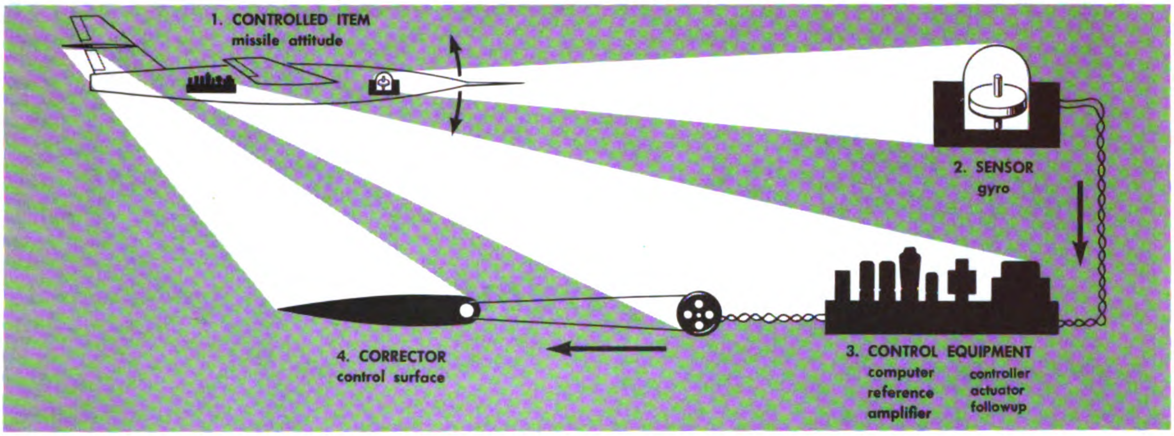
The figure also lists major functions of a guided missile system. Each component forms all or part of the four requirements of a missile system. These major functions are arranged in block diagram form in the figure entitled, General Block Diagram of a missile Control System.

Basic control system units are all applied to one of the eight functions of the basic missile control system.

These functions are not necessarily fixed as shown in the block diagram. Even so, the diagram does provide a basis for understanding component operations.

The control system used in a missile consists of a number of electrical, mechanical, and electronic components which are installed as a part of the missile equipment and are interconnected to operate as a complete control system. The control system is designed primarily to provide stabilization of a missile during its flight from the launch site to the target site. There are also secondary control-system design considerations which are needed because of the extensive research and testing phase a missile must go through before it becomes a recognized tactical weapon. Test range safety requirements also play a part in the control system design of any specific missile.

The control system must be capable of moving the craft's control surface to maintain a given flight attitude and a fixed course or



Four principle parts of every Missile control system

directional heading. The system must be so designed that an operator in a director aircraft or at a ground station can maneuver the missile, through remote control, to accomplish dives, climbs, or various other functions necessary to establish the aerodynamic characteristics of the craft. To accomplish this, the control system must be capable of doing the following:

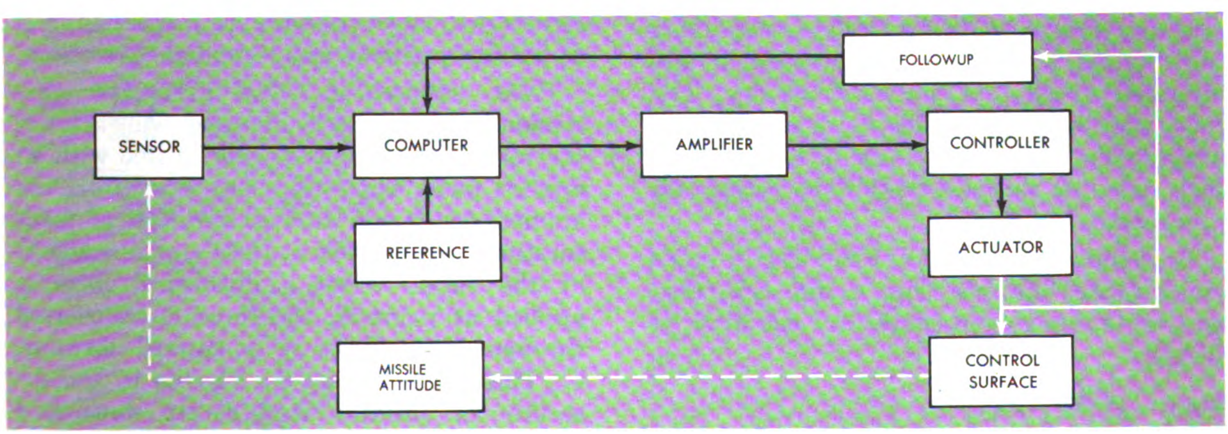
a. It must set up fixed reference lines in space, from which deviations in course or change in flight attitude can be measured.

b. It must provide a mechanical or electrical and mechanical means of operating the control surfaces of the missile as required. Such a device is commonly called a *servo*.

c. It must provide a means for measuring the magnitude and direction of the angular deviations which the missile makes from these reference lines and which the servos make in the position of the control surfaces.

d. It must provide a means for translating the amount and direction of missile deviation and the amount and direction of control surface movement into the operation of a complete system so that the guided missile will perform properly.

The way in which a missile control system meets these requirements is covered in subsequent chapters. In this chapter we will begin our study of the operation of the individual components and subsystems which comprise the makeup of a control system.



General block diagram of missile control system

Sensor Units for Control Systems

A *sensor unit* in a guided missile control system is a device capable of detecting deviation of the missile from a desired flight condition. Sensor units discussed in this chapter include *gyroscopes*, *altimeters*, and *transducers*. Gyroscopes generally are considered to be the basic sensor unit in any missile control system. Altimeters and transducers are used as sensors in secondary or auxiliary servo loops in the control system. Pickoffs, devices that make detected intelligence useful, are also taken up in this chapter.

GYROSCOPES: THE BASIC SENSOR UNITS

Before discussing the application of gyroscopes in a control system, let's first consider some basic gyroscopic terms and definitions.

As stated previously, a *gyroscope* is a mechanical device containing an accurately balanced rotor; the rotor spins about its central or spin axis, which passes through the center of gravity. If it is a *free* gyroscope, it is so mounted that it can tilt or turn in any direction about this center of gravity. The following illustration, Free Gyro, shows such a gyroscope. When the rotor is rotated at a high speed, it assumes the chracteristics of a gyroscope.

It is the characteristic of rigidity in space which makes the gyroscope useful as a reference or sensor unit in controlling the flight of missile.

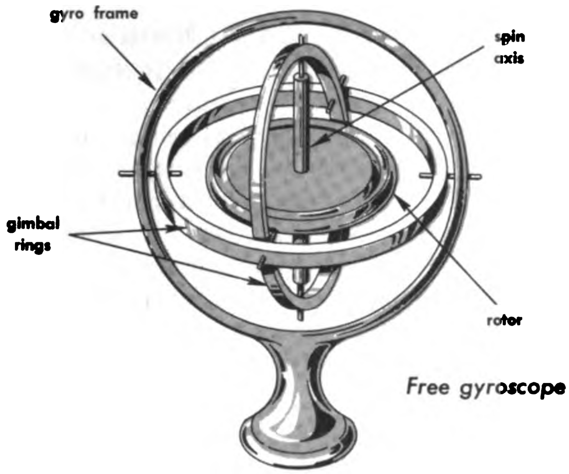
Rigidity of Gyroscopes

Rigidity or *gyroscopic inertia* is that property of a gyroscope which resists any force tending to displace the rotor from its plane of rotation. Three factors determine a gyroscope's strength or amount of rigidity. These factors are: (1) the weight of its rotor, (2) the distribution of this weight, and (3) the speed at which the rotor spins. Rigidity may be increased by adding to the weight of the rotor. A gyro with a heavy rotor has more rigidity than one with a light rotor if the speed of rotation is the same for both. Increased rigidity can be obtained if the weight of the gyro is distributed to the outer rim of the rotor, as far from the spin axis as possible, even though there is no increase in the weight of the rotor. Rigidity is also increased when the speed of rotation is increased. A slowly spinning rotor gives the gyro little or no rigidity.

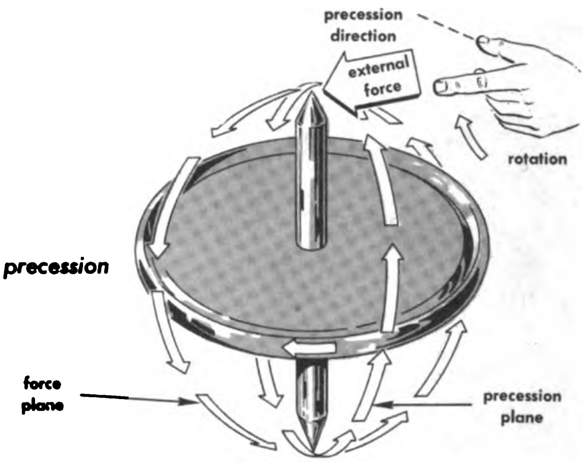
Gyroscopic Precession

Let's consider further the important characteristics of gyroscopic precession. We are concerned with two types of gyro precession: *real* or *induced precession* and *apparent precession*.

REAL PRECESSION. Real gyroscopic precession is that property of a gyro which causes the rotor to be displaced, not in line with an applied force, but 90 degrees away from the applied force, and in the direction of rotor rotation. Take a look at the gyroscope shown at the right. When a downward force is applied at point A, the force is transferred through pivot B as a downward force at C. This force travels 90 degrees in the direction of rotation and causes downward movement at D with subsequent movement (precession) at E.



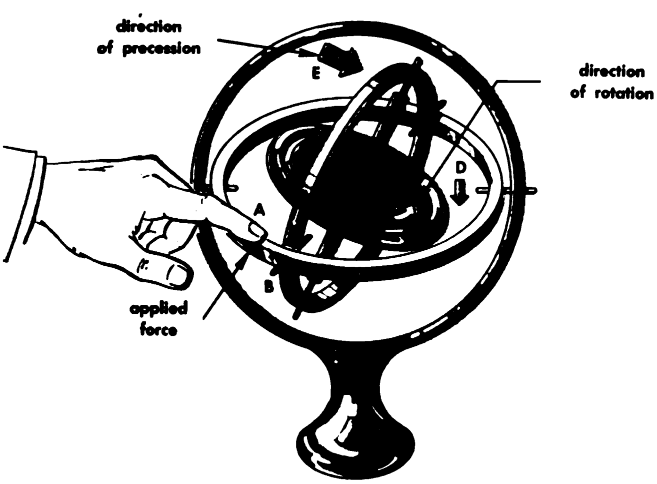
Free gyroscope



Gyro precession

An easy way to remember the direction in which a gyro will precess, when an external force is applied, is by the hand rule as shown in the illustration at the right. Using either hand, place the fingers in the direction of rotation and extend the index finger in the direction of the applied force. The thumb will then extend in the direction of precession.

It is apparent from the preceding explanation of gyro precession that if a force were applied to the gyro at the center of gravity, it would not act to tip the gyro spin axis from its established position. Therefore, no precession would take place. Thus a spinning gyroscope can be moved in any direction as easily as a gyro at rest, provided its axis remains parallel to its original position in space.

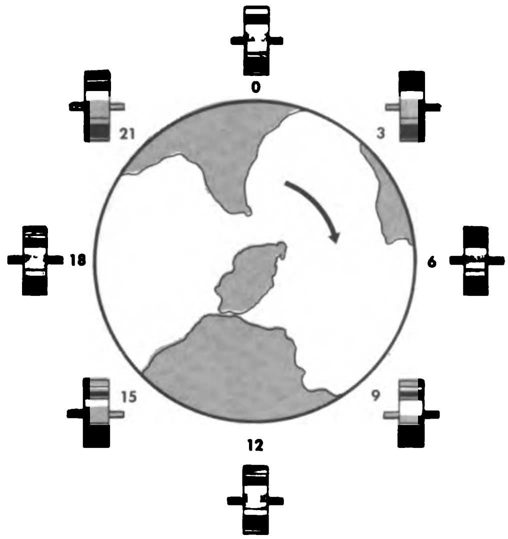


Hand rule for determining direction of gyro precession

The gyro, therefore, provides stability only against tipping its spin axis. Also, a spinning gyroscope can be used to provide stabilization only in planes containing its spin axis. For complete stabilization in every aircraft, two gyroscopes, which have their spin axes at right angles to each other, are required. For this reason, both a vertical gyro and a horizontal gyro are needed to give complete stabilization and to set up the necessary reference lines from which deviation can be measured.

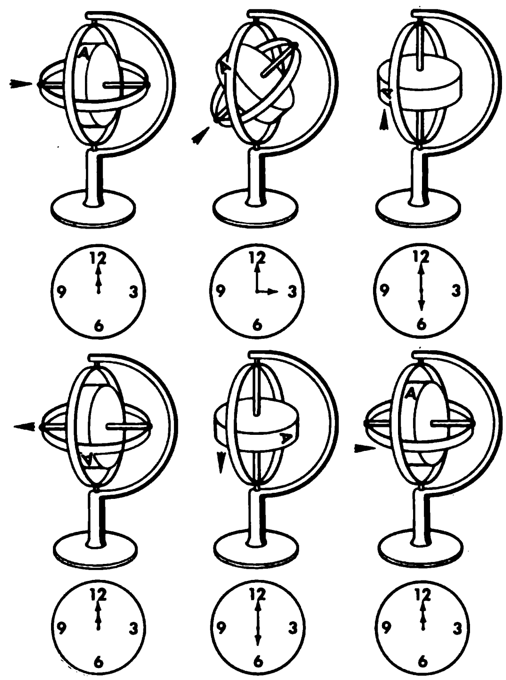
APPARENT PRECESSION. Because rigidity fixes the spin axis of a gyro in space, the axis points in a fixed direction. The earth, which is rotating, turns under the gyro. Thus, the axis of the gyro appears to tilt. For example, imagine a gyroscope at the equator with the spin axis horizontal to the earth and pointed in an east-west direction as shown in the first illustration on page 176. The earth turns in the direction of the arrow (clockwise) with an angular velocity of one revolution every 24 hours. To an observer out in space, the spin axis would appear to maintain its position, pointing east. But to an observer on the earth, the spin axis appears to gradually tilt or drift. At the end of 3 hours, the spin axis has tilted 45 degrees and after 6 hours the spin axis has titled 90 degrees to a vertical position. Notice the sketches of apparent precession on the next page. At the end of 12 hours the spin axis is again horizontal but points west, and at the end of 24 hours it is back in its original position pointing east.

This phenomenon creates the illusion that the gyro has turned over, end for end, and



Gyro position in space about the earth

that a complete revolution is made every 24 hours. Actually, however, the gyro has maintained its position in space, and the earth has moved around it. This movement of the earth in relation to the gyro is called *apparent gyroscopic precession*. The greatest amount of apparent precession is at the equator, and the amount of apparent precession decreases as



Apparent precession

the gyro is moved toward the North or South Pole, at which points apparent precession is zero.

Apparent precession of a gyro makes it unfit for use as a reference over an extended period of time unless some sort of compensating or *erecting mechanism* which keeps the gyro in a fixed relation to the earth's surface is used. Over a relatively short period of time, however, a gyro can be used to establish a satisfactory reference without the use of an erecting mechanism.

Gyro Drift

The line of direction of a gyro is not always in the direction in which it theoretically should point. This error in the gyro is produced by random inaccuracies in the system. The resulting change in position of the spin axis is called *drift*. There are three general sources of drift:

a. *Unbalance*. A gyroscope often becomes dynamically unbalanced when operating at a speed or temperature other than that for which it was designed. Some unbalance exists in any gyro since manufacturing processes do not give perfect symmetry.

b. *Bearing Friction*. Friction in the bearings of the gimbals results in lost energy and incorrect gimbal positions. Friction in the spin axis bearing causes drift only if the friction is not symmetrical. An even amount of friction all around in a bearing results only in a change of the rate of rotation.

c. *Inertia of Gimbals*. Energy is lost whenever a gimbal rotates because of the inertia of the gimbal. The greater the mass of the gimbal, the greater the drift from this source.

The complete elimination of drift in gyroscopes appears to be an impossibility. However, great strides have been made in recent years toward reducing the amount of drift. The methods being used are discussed briefly later in this section.

Gimbal Lock

If two gimbals of a gyro are positioned in the same plane, the gyro is not free to precess; as a result, the forces of precession lock the gyro in a rigid position when a torque is applied.

The two basic properties of a gyro, rigidity in space and precession, are utilized in gyroscopic instruments. Rigidity is utilized to establish a reference in space unaffected by movement of the supporting body; and precession is utilized to control the effects of the earth's rotation, bearing friction, and unbalance, thus maintaining the reference in the required position.

Rate Determination

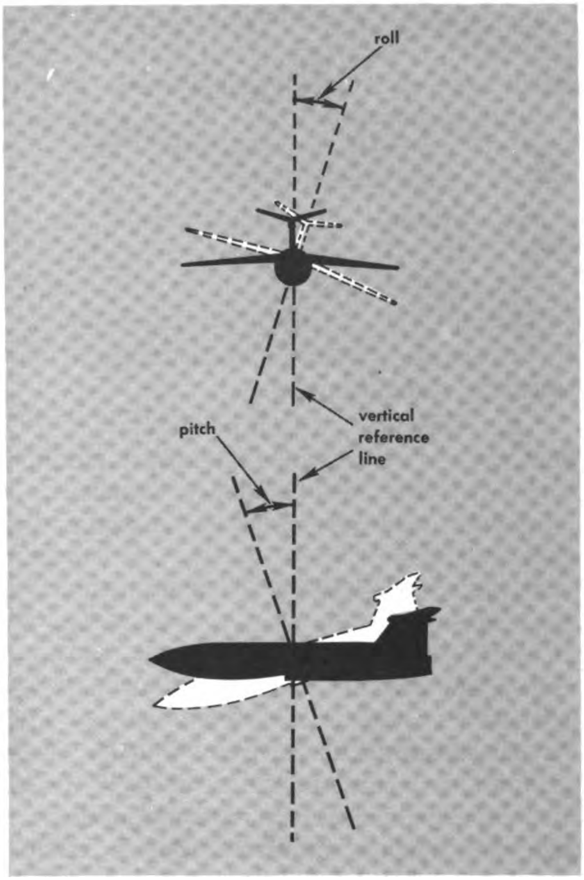
When in a missile, a gyroscope with two gimbals, one of which is restrained by springs or a similar means, exerts a force on the restraint which is proportional to the angular rate of movement of the missile in one plane. Thus, by measuring the force exerted on the restraint, an indication of rate of deviation is obtained directly.

Use of Gyroscopes in Guided Missiles

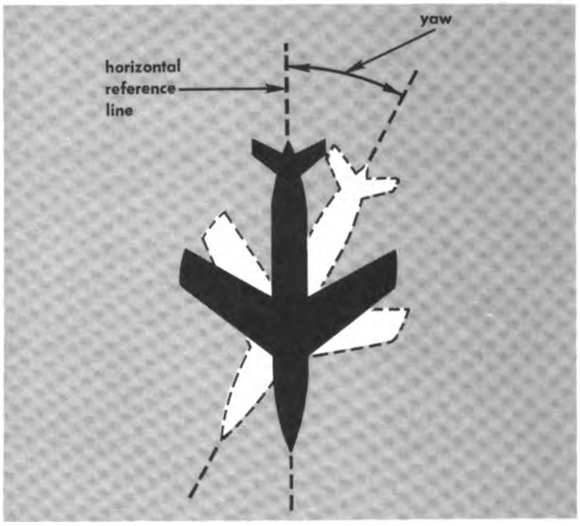
A minimum of two *displacement sensing* gyros is necessary missile stabilization. Each of the two gyros is used to set up a fixed reference line from which missile deviation, either in directional heading or change in flight attitude, is measured. One of these lines is vertical and is established perpendicular to the earth's surface. This line passes through the spin axis of a vertical gyroscope. From this line, deviations can be measured from flight attitude about the roll axis or about the pitch axis of the missile, as shown in the illustration at the right above.

Deviations from directional headings cannot be measured by a vertical reference line. A second reference line fixed in space horizontally must be set up so that deviations in directional heading can be measured. This reference line passes through the spin axis of a horizontal gyro and is, therefore, stabilized in space. Since deviations from course are measured from this reference line, it is set up parallel to the horizontal axis of the missile as shown in the accompanying illustration of a horizontal reference line.

Gyros used for missile applications are divided into two classes: (1) gyros used for stabilizing purposes, i.e., control system gyros, and (2) gyros used for both guidance and stabilization.



Vertical reference line showing deviation about roll and pitch axis



Horizontal reference line showing deviation about yaw axis

Gyros used for stabilizing purposes alone are usually adaptations of existing gyro structures. A typical *displacement sensing* vertical or horizontal gyro to be used with external guidance signals might have the following characteristics:

a. Power — alternating current, two or three phase, 400-2000 cycles per second.

b. Sensing element — magnetic or capacitive pickoff, or precision potentiometer.

c. Size — 5 in. to 6 in. cube, in hermetically sealed housing.

d. Weight — 3 to 5 pounds.

Such a gyro provides adequate missile stability when supplied with guidance signals at a precession rate greater than 1 degree per minute.

As shown previously, a minimum of two displacement sensing gyros are necessary for missile stabilization. When turns or other maneuvers are indicated, it often becomes necessary to use a three-gyro system with but one sensing axis per gyro. Usually three-gyro systems can use the same gyro design in all three positions, thus reducing production and maintenance problems. For a specific application, it may be possible to use a displacement sensing gyro as small as 4 x 4 x 4 inches over-all, weighing less than 3 pounds.

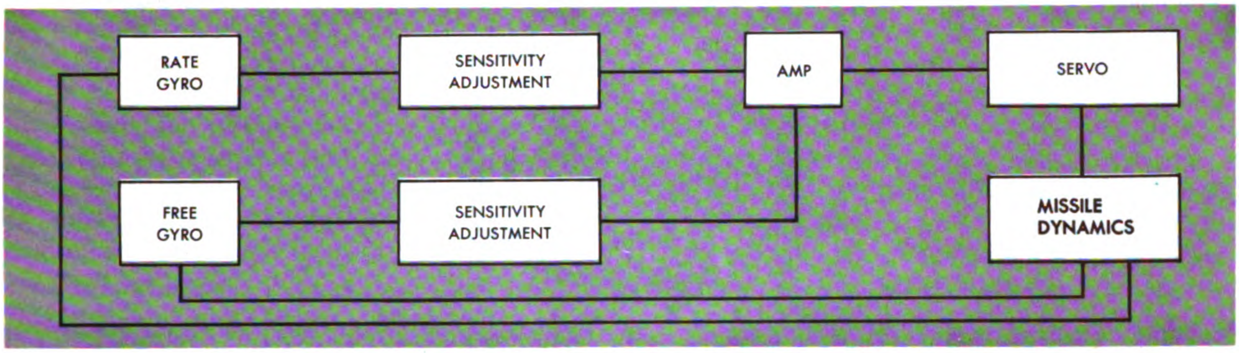
Gyros used for both guidance and stability are free gyros which depend upon design skill and accurate manufacturing techniques to provide the necessary space reference. As a class, such gyros are usually larger and heavier than nonguidance types. In general a *guidance* gyro is built around a comparatively large, high-speed rotor of tremendous angular momentum, supported on micro-friction gimbal bearings. A "perfect" free gyro of this type, balanced in neutral equilibrium, would maintain its spin axis in a constant angular position in space. Two or three such reference axes, usually mutually perpendicular, can be used as the basis of a guidance system. Guidance gyros tax the skill of the designer-manufacturer to the utmost.

Rate Gyro

In addition to the control signals supplied by the vertical and horizontal gyros, which are proportional the deviation of the missile, a signal proportional to the rate of deviation is required for accurate control and smooth recoveries. This *rate-of-deviation* signal is supplied by a rate gyro. A rate gyro has a restricted gimbal, free to rotate about one axis only. It is, however, a true gyroscope, conforming otherwise to the same basic principles as other gyros. The spin axis of a yaw rate gyro is mounted with its spin axis parallel to the missile line of flight. A roll rate gyro is mounted with its spin axis parallel to the missile pitch axis, at right angles to the line of flight. A pitch rate gyro is mounted with its spin axis parallel to the yaw axis of the missile, also at right angles to the line of flight.

RATE GYRO APPLICATION. The use of displacement signals alone as a means of applying corrective control in an autopilot control system results in a tendency of the craft to yaw or pitch about its desired course. When displacement signals are used alone, the application of control to bring the craft back on course results in overcorrection because of the momentum of the missile. The control system, therefore, should have not only a means of recognizing the position of the craft in space, but also a means of detecting the rate of position change. By adding the rate of position change signal to the displacement signal, the tendency of the craft to overcorrect is minimized and a better degree of stability is obtained.

Present systems of missile stabilization employ both free and rate gyros, or a differentiating network in conjunction with a free gyro, in order to provide both position change and rate of position change signals to the control servos. The introduction of the rate signals results in good aerodynamic damping. The figure above right shows a functional diagram of a system being used to sense motion of a missile about its control axis. The system is for a missile control channel using both a rate gyro and free gyro. The sensitivity adjustments are used to set the ratio of the rate signal from the rate gyro to the position signal from the free gyro. These signals are fed to an amplifier which adds the two signals and gives an output voltage proportional to their sum. This voltage is



Missile control channel employing a rate gyro and a free gyro

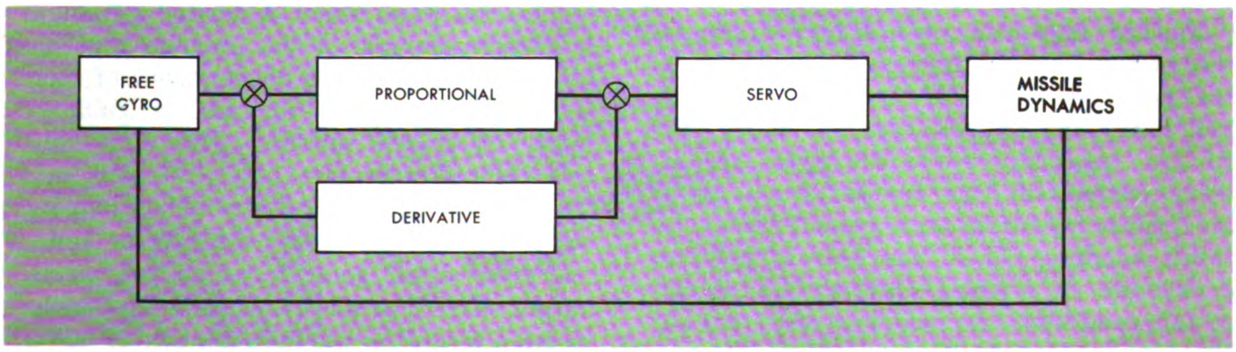
applied to the servo which positions the control surface in the proper position to drive the angle of change and rate of change to zero.

Shown below is a functional block diagram of a control channel employing a free gyro and a computer network. The voltage signal from the free gyro goes to the proportional and derivative channels of the computer. The proportional channel consists merely of gain control. In the derivative channel the signal is differentiated by an R-C network and amplified. These two signals are then combined and fed to the servo. The servo positions the control surface, and the angle of deviation is driven toward zero by the response of the missile to control surface deflection.

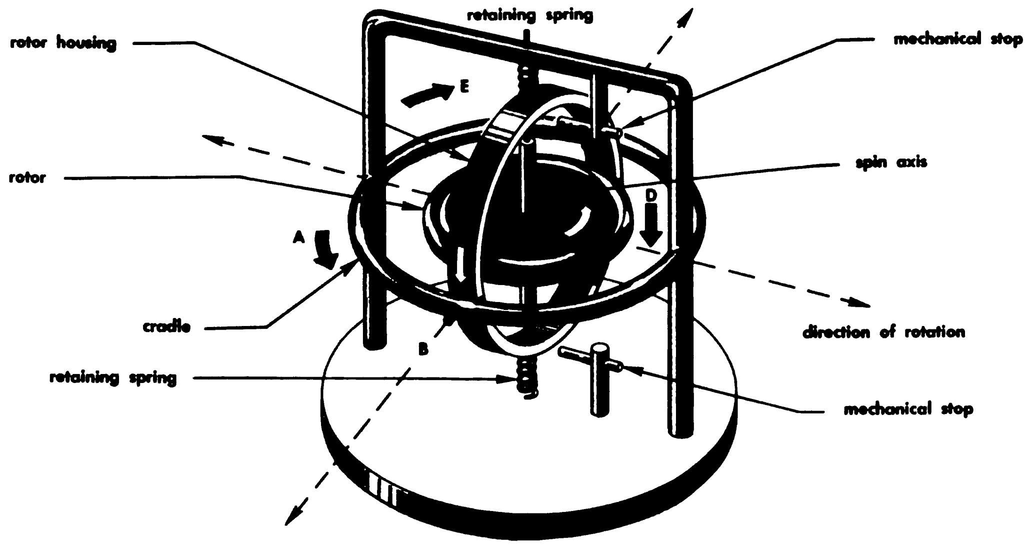
The rate gyro used in the first application is similar to a free gyro in that it has a high gyroscopic momentum while the rotor is spinning. It differs, however, in that it is free to rotate about one axis only and in that it is sensitive only to angular rates of movement; that is, angular movement of its cradle in excess of a certain rate causes its spin axis to change its position in relation to space. Such changes occur 90° in the direction of rotation from the angular movement applied, and in the direction of movement. The illustration on the next page demonstrates the action of a rate gyro when used in a missile control system.

An angular force applied to the cradle at "A" is transmitted through pivot "B" as a downward force at "C." This force travels 90° in the direction of rotation and causes downward movement at "D" with subsequent precession at "E." As soon as the angular rate applied to the cradle ceases, restraining springs restore the gyro to its neutral position. Stops are provided to limit the precession of the gyro to a few degrees in either direction.

Rate gyros may prove to be of considerable value in high-performance missile applications.



Missile control channel employing a free gyro and computer network



Action of rate gyro

Three rate gyros, combined with compact electronic integrators, offer an extra good source of displacement reference whenever guidance is not a function of the gyro system. Control systems of this type are within the scope of gyro and electronic experience

Gyroscope Error

As you know, random drift in a gyroscope is that drift caused by bearing friction and dynamic unbalance. This type of drift results in an unpredictable precession of the gyro. Apparent drift in a gyro is that drift which occurs because of the rotation of the earth. It is toward the elimination of random drift that designers and manufacturers of gyros are now concentrating their utmost efforts. Until a few years ago, the most accurate gyros available had random errors of plus or minus 1 minute of arc per minute of time, or 1 degree per hour.

Marked improvement has been made in the performance of free and rate gyros in the past few years. However, these units still have high uncertainties due mainly to friction, especially when the gyro case is being accelerated. The present types of gyros are delicate instruments and cannot withstand rough handling. Their fine bearing surfaces deteriorate in storage. Nevertheless, under certain conditions of missile launching and for the simplest system of stabilization, the use of the free gyro is often required.

One of the requisites for an ideal missile guidance system specifies that the system be invulnerable to enemy jamming. To date, only two of the many proposed guidance systems, the celestial system and the inertial system, meet this requirement. Both of these guidance systems depend upon gyros for their basic reference. The celestial system must have a gyro stabilized platform for its telescopes, and the inertial system must have a stabilized platform for its accelerometers. Random errors of the order of one degree of arc per hour of time would result in a missile error of 60 nautical miles per hour of flight. Thus, even for a flight of only a few minutes duration, the expected accuracy is not great.

Since the main cause of random drift in gyros is gimbal bearing friction, the problem of improving gyro stability has been approached with a view to reducing this friction.

FLOATED GYRO UNIT. A floated gyro unit is a good example of the progress that has been made toward the development of more accurate instruments for use as guidance gyros in guided missiles. This unit, also called the *Draper gyro* and *HIG gyro*, is a viscous-damped, single-degree-of-freedom gyro with a microsyn torque generator and a microsyn signal generator mounted on its output shaft. The miscrosyn torque generator places a torque on the gyro gimbal.

A single-degree-of-freedom gyro is so called

because its gimbal is free to rotate with respect to the gyro case about a single axis. This axis, called the output axis, is perpendicular to the gyro spin (reference) axis. The third axis, the gyro input axis, is perpendicular to both the spin axis and the output axis. If the gyro case is subjected to an angular velocity with a component about the input axis, a precessional torque develops about the output axis.

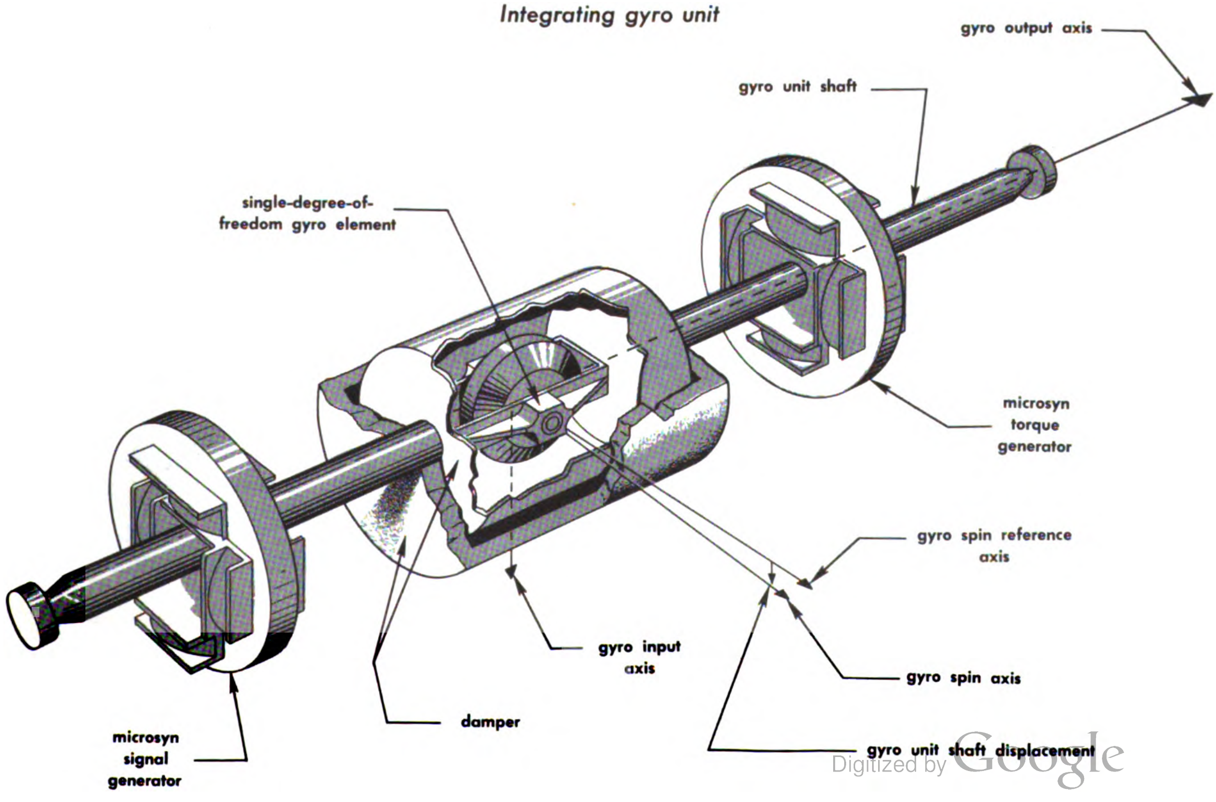
The illustration below shows a cutaway view of the integrating gyro unit. The gyro wheel is contained within the damper housing. The microsyn signal generator and torque generator units are mounted on the gyro shaft as shown. The space between the damper housing and the gyro case is filled with the viscous damping fluid of high specific gravity. Because of the high specific gravity of the fluid, it serves to float the gyro damper housing and gyro gimbal shaft, thus reducing the gimbal bearing friction. In this way, the random drift due to friction is greatly reduced. Because the gyro is supported by the viscous damping fluid, the unit is relatively free from the undesirable acceleration effects experienced with the usual gyro construction. This is a decided advantage of the integrating gyro for application to guided missiles.

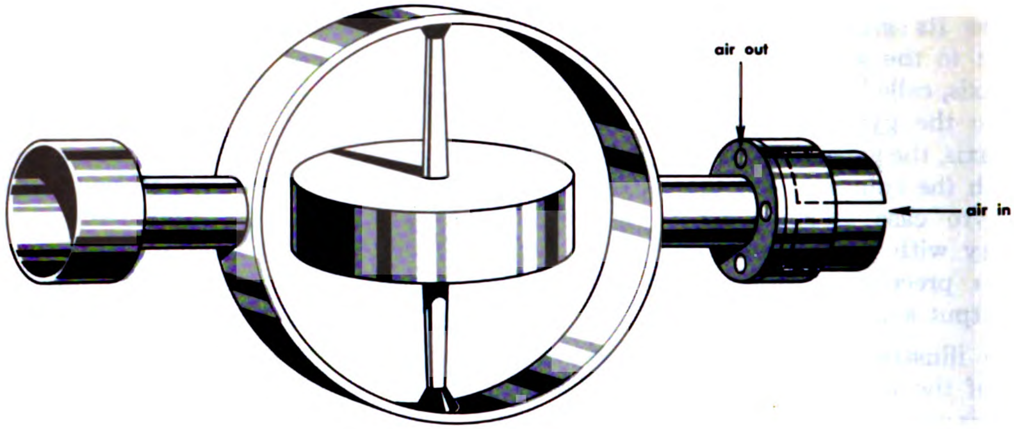
The sensitivity of the integrating damper for angular velocity input and torque output is a function of the viscosity of the damping fluid, which in turn is dependent on the temperature of the fluid. Consequently, heaters with suitable thermostatic controls are placed in the gyro case around the space containing the damping fluid to keep the fluid at the desired operating temperature, thus maintaining the desired damping characteristics.

If this gyro is mounted in a missile so that its input axis is parallel to the pitch or yaw axis of the missile, the torque applied to the output shaft would be proportional to the difference between the desired angular velocity of the missile and its actual angular velocity about the respective axis. Since the liquid has the function of integrating the torque, the unit can be considered as a combination angular velocity error measuring instrument and an integrator in cascade.

AIR BEARING GYROS. Another method by which random gyro error due to bearing friction has been greatly reduced is through the use of air bearings. With this type of bearing, friction is reduced to a degree so negligible it might be considered a zero component. An air bearing works on the same principle as a vacuum cleaner display in which a large rubber ball is held suspended

Integrating gyro unit





Air bearing gyro

in a conical air stream. The basic principle of an air bearing gyro is illustrated in the above diagram.

Air bearings thus far developed have exhibited an unpredictable *auto-rotational* or turbine torque caused by a symmetrical air flow and eddy currents. Continued development doubtlessly will eliminate this factor, since it is strictly a design consideration. Initial successes in this direction indicate that a hundredfold improvement is possible in gyro stability.

Gyros: Heart of the Missile Control System

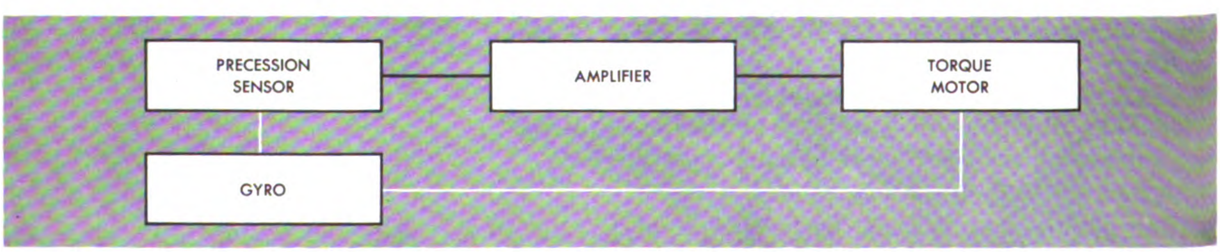
The gyro system is the heart of the missile control system. The function of a gyro system is ultimate actuation of the aircraft controls to maintain the missile in a specified attitude.

Accuracy requirements for control system gyros are not so rigorous as for guidance system gyros. Gyroscopes of World War II vintage still are used in the control systems of many missiles. On first thought, one might not readily see why the displacement gyro used in a missile autopilot need not be as accurate as a gyro used in a missile guidance system. However, closer examination reveals that entirely different principles are envolved. In the case of a guidance gyro, it is necessary to have an instrument which gives an accurate reference of the missile relative to its desired trajectory on the earth. In the case of the displacement gyro, the instrument must furnish a reference to the missile control system relative to its flight attitude only.

A vertical displacement gyro can be equipped with an erection system to keep it in a relatively level position. The erection system must be capable of maintaining the displacement gyro within allowable limits necessary for flight attitude references.

GYRO ERECTION AND SLAVING CIRCUITS

Displacement gyros are controlled or kept erect by means of an erection and/or slaving system. Vertical gyros are controlled by an erection system, while horizontal gyros are controlled by both a slaving system and an erection or leveling system.



Vertical gyro erection system

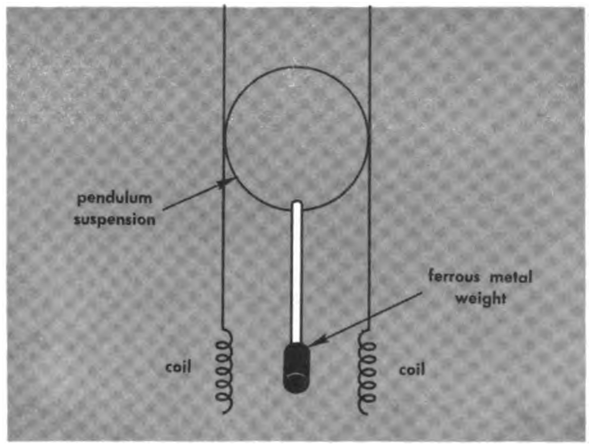
Vertical Gyro Erection

A vertical gyro is used to stabilize a missile about the pitch and roll axes. Two torque motors are provided to maintain the gyro axis erect in both planes. A basic vertical gyro erection system is represented to the left below in block diagram form.

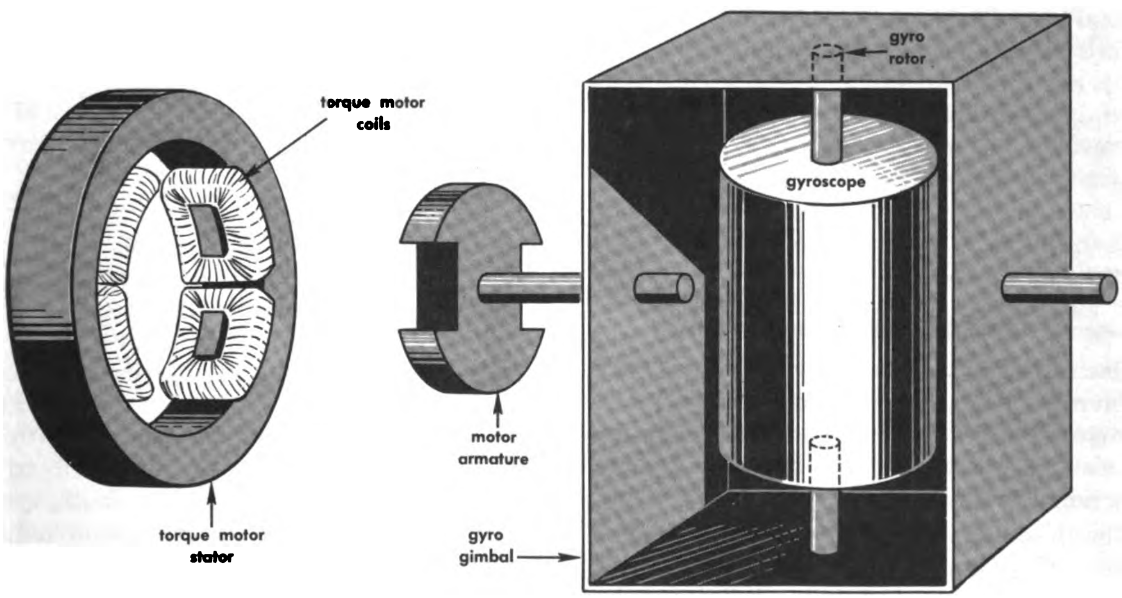
The gyro precession sensing device (pick-off) may be any one of several types. For purposes of explanation, the pendulous weight type shown on the right is discussed here in detail.

Four secondary windings are connected in pairs, each coil in a pair being phase-opposed. The pairs are located beneath a pendulous, ferrous metal weight which acts as the magnetic coupling between the primary winding and the four secondary windings. When the gyro axis is vertical, the coupling to each pair of secondary windings is equal, but since the coils of each pair are phase-opposed, the signal output is zero.

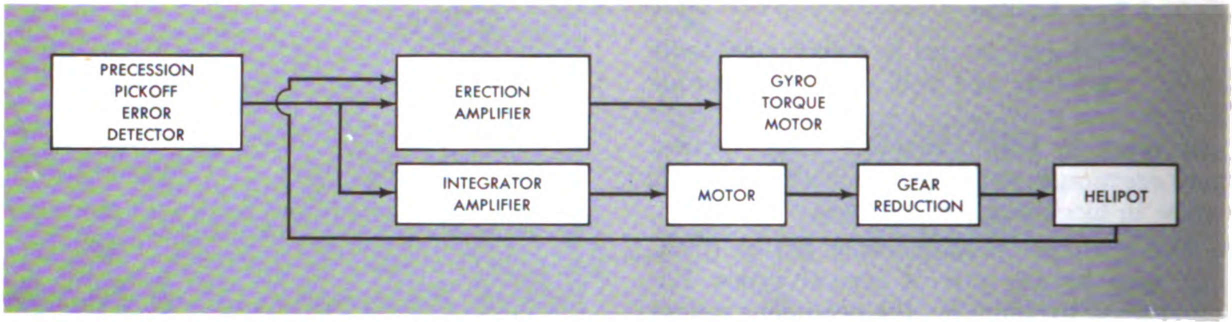
The coils of this transformer-like device are wound on the gyro rotor, and the pendulous weight is suspended in the shaft of the rotor. Therefore, when the spin axis of the gyro moves away from the vertical, the pendulous weight increases the coupling to one coil of a pair and decreases the coupling to the other. The amplitude of the signal thus developed is determined by the amount of gyro precession, and the phase of the signal depends upon the direction of precession. The signal, which may appear across one or both pairs of secondaries, is then amplified and sent to the appropriate torque motor, which applies a force to cause the gryo spin axis to precess back to the vertical position. The pendulous weight is liquid damped to prevent excessive oscillations.



Gyro precession pickoff



Gyro and torque-motor assembly



Gyro erection system with integrator

The *erection amplifier* is a conventional amplifier capable of increasing the precession signal to a voltage large enough to drive the torque motor.

The *torque motor* is generally a two-phase, squirrel-cage type induction motor. Some erection systems may include an integrator loop to correct for any dynamic unbalance in the gyroscope. A block diagram of such a system is shown above. The precession pickoff feeds a signal to the erection amplifier which amplifies it to a large enough voltage to drive the gyro torque motor. The integrator amplifier also receives the signal from the precession pickoff, and its output is applied to a motor which drives the wiper arm on a helipot through a large gear reduction. Thus, for occasional small errors, the helipot output is very small, but for a constant error, which would result if the gyro were dynamically unbalanced, the helipot is driven long enough to completely compensate for the gyro unbalance. The helipot output is fed back to the erection amplifier input and appears as a constant signal to the gyro torque motor.

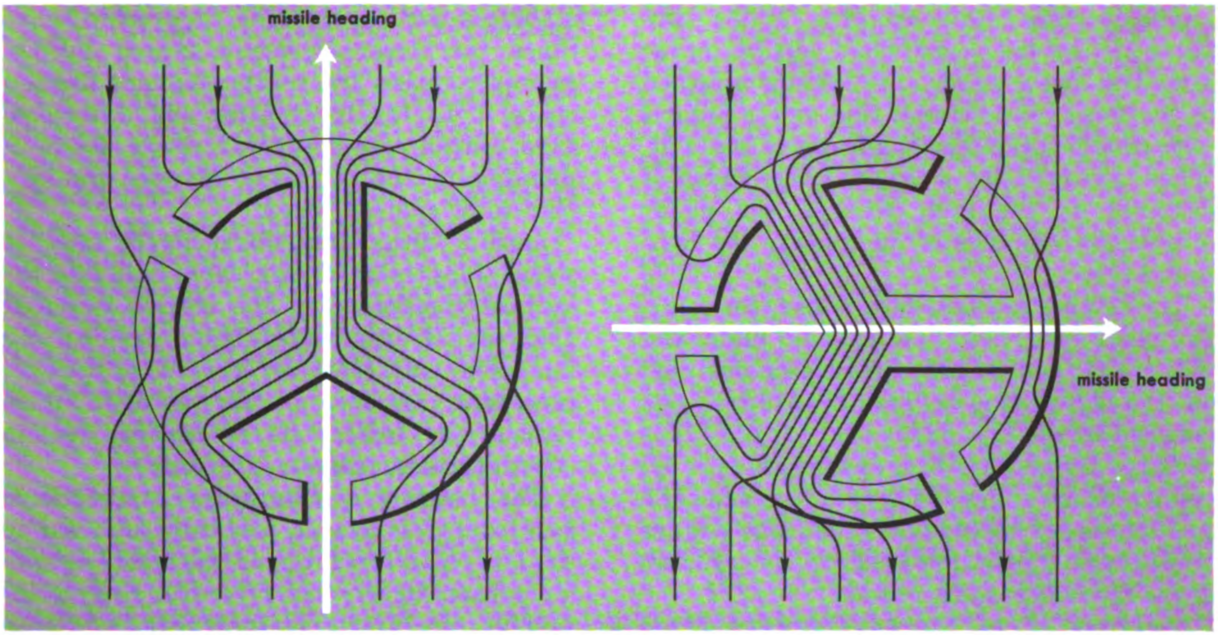
Gyroscope Slaving and Erection Systems

Horizontal gyros use a leveling system and a slaving system. The leveling system serves to maintain the gyro on a level plane, while the slaving system permits the gyro to give a directional indication.

The best known reference for slaving a horizontal gyro is the earth's magnetic field. This is accomplished by use of a flux valve which is a unit that senses the direction of the earth's magnetic field. It consists of a primary coil and three secondary coils wound on a metal core. Power to energize its primary coil and the signals originating in its secondary coils are transmitted through terminals located under a compensator which serves as a cover. The flux valve unit is pendulously suspended on a universal joint within the bowl and is weighted so that, within limits, it continuously responds to gravity. To prevent excessive swinging in flight, the bowl is filled with a damping fluid.

The fundamental flux valve unit is mainly a spider-like core made of laminated metal of high permeability. The core resembles a three-spoke wheel, slit through the rim between the spokes. Note the illustration on the right. Its hub is widened to receive a core about which the primary (exciter) coil is wound. The coil is energized by a 23-volt, 400-cycle power source. Each spider leg is encircled by a pickup (secondary) coil and terminates in arcs of metal which serve as flux collectors.

The flux valve unit functions in this manner: As the positive half-cycle of the exciting current builds up, it sets up a magnetic field which expels the earth's flux from the spider legs. The expelled flux cuts the pickup coils on each leg, inducing current in each. As the positive half-cycle of the exciting current decays, the earth's flux, because of the higher permeability of the legs compared to the surrounding air, is drawn into the legs, again cutting the pickup coils, and again inducing current but of opposite polarity. Thus, a half-cycle of alternating current produces one full cycle of induced current. The negative half-cycle of the exciting alternating current



Changes in flux through flux valve as heading changes

reproduces the action of the positive half-cycle. Since the exciting current has a frequency of 400 cycles per second, the induced current is doubled in frequency to 800 cycles per second.

The magnitude of the induced current in the pickup coils varies according to the number of magnetic lines cutting them. And the number of lines cutting the coils varies according to the position of the individual spider leg with regard to the magnetic north. Thus, each pickup coil produces a voltage of a given magnitude, which is transmitted to the corresponding "Y" coils of the flux valve synchro in the directional gyro control. The magnitude of the transmitted voltage is dependent upon the pickup coil's position relative to the magnetic meridian. This transmission creates a magnetic vector. The vector varies in its position across the stator in the same relationship as the spider legs of the flux valve unit vary in their orientation to the earth's magnetic lines. The rotor of the flux valve synchro, if positioned at right angles to this vector, produces no voltage. For the purposes of this discussion, such a no-voltage condition will be referred to as rotor alignment with the vector.

The compensator counteracts constant distortions originating from magnetic aircraft parts and electrical apparatus located in the vicinity of the flux valve. The compensator is a flat, compact unit which contains four permanent magnets to oppose the deflecting forces. The magnets are rotatable by means of two slotted shafts. When the compensator is attached to the transmitter (flux valve), the shafts are clearly identified. One shaft has "NS" stamped adjacent to it and the other "EW." Alignment of the dots on the shafts with those on the compensator cover indicates that the magnets are in neutral position.

Type S-1 Directional Gyro Control

A Type S-1 directional gyro control houses an electrically driven gyro. The gryo's spin axis is not only maintained tangent to the earth's surface, but is also *slaved* to the earth's magnetic field for the purpose of furnishing a basic magnetic reference to which the magnetic heading of the missile can be slaved.

Most of the internal assemblies are supported by a frame which is mounted on a cast base. The cover, which is transparent to permit reading of a compass card, fits over all of the internal assemblies and fastens onto the base.

The leveling torque motor is of the two-phase induction type and has a vertical torque axis. The motor consists of a stator attached to the top of the vertical ring and a squirrel cage (rotor) secured to the frame of the unit. When actuated by a liquid level, which is mounted on the gyro housing, the motor returns the spin axis of the gyro to the horizontal plane.

The slaving torque motor is likewise of the two-phase induction-type. It has a horizontal torque axis and consists of a stator mounted on the gyro housing and a squirrel cage fastened to the side of the vertical ring. It is actuated by an amplified and phase-detected voltage which aligns the gyro to the magnetic heading sensed by the flux valve. The voltage originates in the rotor of the flux valve synchro in the directional gyro control.

The rotor of the flux valve synchro of the directional gyro control is fixed to the lower pivot of the vertical ring and turns relative to its stator, which is fastened to the frame of the unit. When the switches are thrown to energize the system, the transmitter instantly establishes a magnetic vector across the stator of the flux valve synchro. This magnetic vector is the resultant of the earth's magnetic lines of force as detected by the three transmitter spider legs. The flux valve synchro rotor, fixed to the vertical ring or outer gimbal of the gyro, may not be aligned with that vector. Any misalignment between the rotor and the vector is itself the corrective factor, for under this condition voltage is, and will be, continuously generated in the rotor. This voltage, after amplification and phase detection, energizes the slaving torque motor until perfect alignment between the synchro rotor and vector is attained. The gyro rotates the vertical ring as it responds to the precessing force of the torque motor. Such rotation brings the rotor of the flux valve synchro into alignment with the vector, stopping generation of voltage in the rotor and halting the action of the torque motor. Should the spin axis of the gyro drift in azimuth during flight, a similar action takes place. Conversely, effective alignment between the synchro rotor and the vector is maintained during turns by the rotation of the directional gyro control case, with its fixed synchro stator, about the gyro-stabilized vertical ring and its synchro rotor. Since the magnetic vector sensed by the remote compass transmitter remains fixed in space, both the transmitter and directional gyro control case effectively rotate about it, maintaining the alignment between the synchro rotor and the vector.

During turns, centrifugal forces tend to swing the pendulous flux valve unit into the vertical component of the earth's magnetic field, thus slightly distorting the signals of the flux valve unit. However, because of the slow precessing response of the gyro to these signals, the indications of heading remain effectively accurate.

The heading synchro of the directional gyro control is located vertically adjacent to the flux valve synchro, and the rotors of both are affixed to the same vertical ring pivot. The heading synchro stator is attached to the frame above the flux valve synchro stator. The rotors or stators of both synchros move simultaneously, depending on whether the vertical ring or the case produces the rotation. The heading synchro rotor is energized by a 115-volt, 400-cycle power supply to produce a magnetic field which induces voltages in the "Y" connected coils of its stator.

When the missile turns in either direction, the flux valve, being fixed to the missile, turns with it. As it turns, the flux valve spider legs and coils also turn and continuously change positions and relationships relative to the magnetic meridian. Because of such changing relationships, the earth's flux penetrates each spider leg in differing quantities. Consequently, the magnitude of the induced voltages in each spider pickup coil continuously varies. Correspondingly, it varies in the "Y" connected coils of the stator of the flux valve synchro in the directional gyro control. Thus, new vectors are being continuously formed across both the transmitter and flux valve synchro stator coils which, in effect, rotate relative to these coils as a pointer would move over the face of a dial. However, the directional gyro control case, to which the synchro stator is fixed, also rotates with the missile and angularly displaces the stator, Therefore, these magnetic vectors, although rotating with respect to the stator coils, remain constantly aligned with the flux valve synchro

rotor which is held rigidly fixed in position by the gyro. Under such conditions, misalignment could occur only should the gyro drift.

The slaving amplifier used performs two functions: first, it amplifies the signal of the rotor of the flux valve synchro in the directional gyro control; second, it detects its phase, thus controlling the direction and amount of torque in the slaving torque motor. The diagram below shows a J-2 slaving system.

We will now consider the altimeter, which is the second type of sensory unit covered in this chapter.

ALTIMETERS: SENSOR UNITS FOR MEASURING ALTITUDE

Instruments used to measure altitude are called altimeters. Two main types of altimeters are *pressure altimeters* which give an approximate true altitude from which a more accurate value can be calculated, and *absolute altimeters* or *radio altimeters* which give absolute altitude directly.

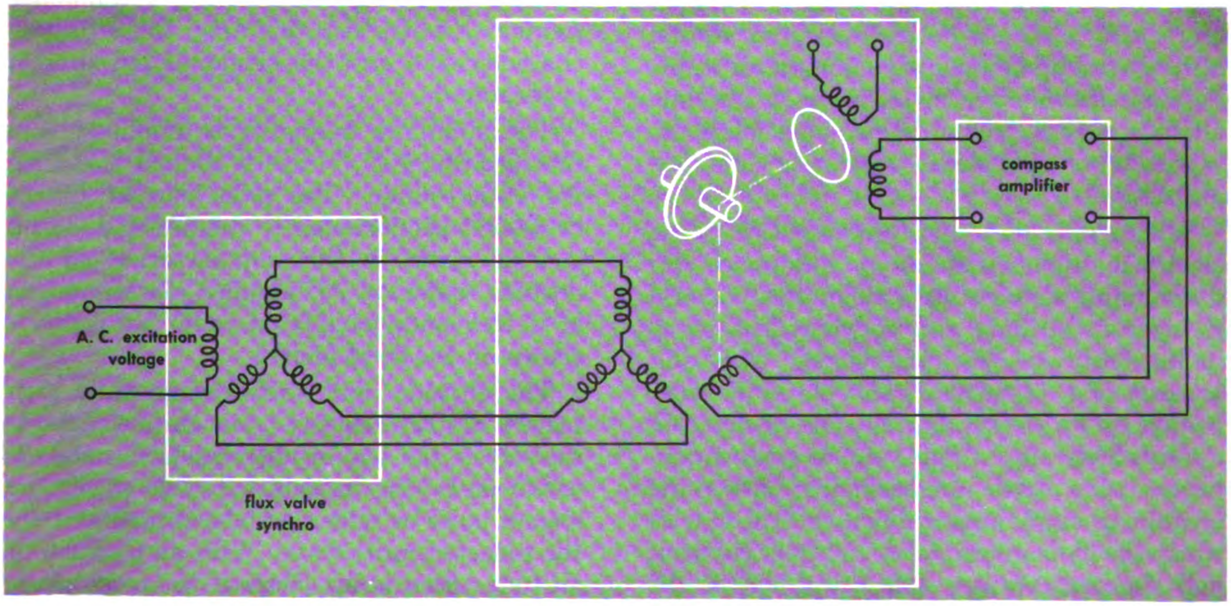
Altitude refers to the height of an aircraft in the air. Altitude is defined as vertical distance above some point or plane used as a reference. It follows, then, that there may be as many kinds of altitude as there are reference planes from which to measure.

Pressure Altimeter

A pressure altimeter is simply a mechanical aneroid barometer, registering atmospheric pressure on a scale calibrated in terms of altitude instead of inches of mercury. It consists of a small, airtight chamber from which most of the air has been removed. The pressure or weight of the outside air tends to collapse the chamber, but this tendency is resisted by a spring. As the atmospheric pressure increases, the chamber is compressed; as it decreases, it is again expanded by the spring. This slight motion is magnified mechanically, and it is registered in terms of the altitude that would produce a corresponding change in pressure under standard conditions.

The pressure altimeter although calibrated in feet, actually measures atmospheric pressure at flight level and interprets this value in terms of feet above a certain pressure level. If the pressure were constant for each level of altitude, an altimeter could be designed to indicate the true altitude corresponding to each pressure. But since the pressure does not remain exactly constant at any one level, the altimeter cannot indicate true altitude directly.

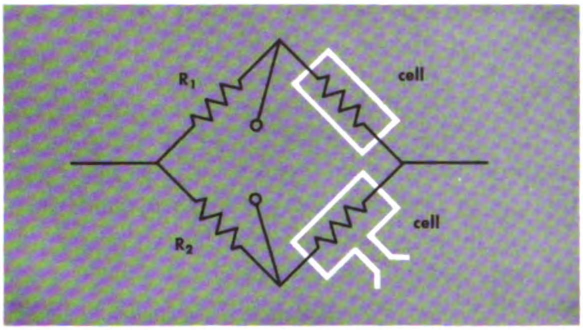
As a compromise, the altimeter is designed to indicate changes of altitude as the pressure varies according to an arbitrary rate. This



Basic functional diagram of J-2 slaving system

arbitrary rate is called the standard pressure lapse rate. For any change of pressure, the altimeter indicates the corresponding change of altitude according to the standard pressure lapse rate.

The pressure altimeter is designed to measure altitude from 28.00 in. to 31.00 in. of mercury above any one pressure level. Whatever pressure is set in the window on the dial face is the pressure level above which the altimeter measures standard altitude.



Altimeter cells in bridge network

Accurate absolute altitude is an important requisite for good navigation and bombing. Absolute altitude can be computed from the pressure altimeter readings, but the results are often unreliable. Under changing atmospheric conditions, corrections applied to obtain true altitude are only approximate. Besides, any error made in determining the terrain altitude results in a corresponding error in the absolute altitude. Such an error in absolute altitude makes ground speed by timing unsatisfactory for dead reckoning at low altitudes.

ALTIMETER CELL. An *altimeter cell* is an electrical application of a barometer for determining altitude. An altimeter cell detects changes in altitude and converts this information into an electrical signal. The unit consists of a filament of fine platinum wire, heated by an electric current and enclosed in a vented envelope. The vent is always connected to a static pressure line. When the altitude or static pressure changes, the rate at which the filament can release its heat changes, and the temperature and resistance of the filament also change. This characteristic is utilized by connecting the filament as an arm in a Wheatstone bridge. Usually two cells are placed in the bridge as shown on the left. One cell is vented and one sealed, in order to compensate for surrounding temperature changes. The signal output of the bridge is proportional only to pressure changes since both cells change with temperature while only one changes with pressure.

An altimeter cell measures altitudes as high as 500,000 feet. This wide range gives the cell an advantage over a mechanical aneroid. Another advantage of altimeter cells over a mechanical aneroid is that a unit consisting of two cells is quite compact, the height of each individual cell being only 1½ inches.

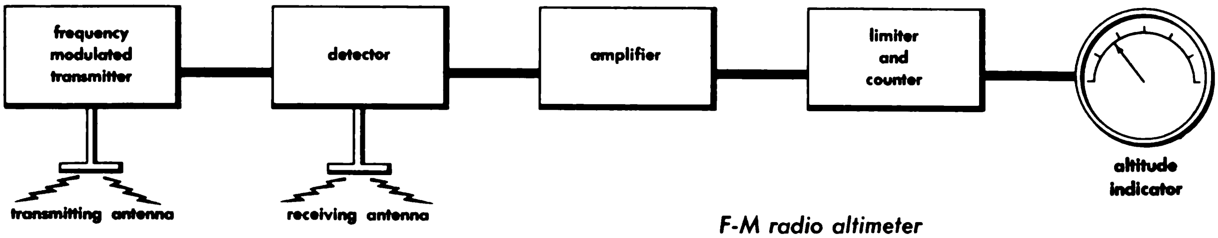
Radio Altimeters

Frequency-modulated radio altimeters have been in common use on many types of aircraft for some time. These altimeters can be used to automatically maintain a missile at a preset altitude. It is also possible to use the radio altimeter as a proximity detonating device for large missiles.

In frequency-modulated radio altimeters, a transmitter radiates toward the earth a wave which is frequency-modulated by a symmetrical triangular modulating voltage. The reflected energy is received on a separate antenna and is combined with energy taken directly from the transmitter. A difference frequency which depends upon the altimeter characteristics and the height of the aircraft is generated.

The difference frequency is amplified, and the resulting AC voltage is limited and then fed to a counter. The plate current of the counter is proportional to the altitude, and a current meter in this circuit serves as the altitude indicator. A simplified block diagram of a frequency-modulated radio altimeter is shown on the right above.

The altitude indicator on some radio altimeters is a cathode ray tube. On the cathode ray screen, the zero or reference lobe (pulse) is produced by the transmitted radio wave; the second pulse or reflection lobe is produced by the reflected radio wave. With an increase of altitude there is an increase in travel time, which results in an increase in distance between the two lobes displayed on the CRT



F-M radio altimeter

screen. Thus a proportionate scale can be constructed on the face of the cathode ray screen which indicates the altitude by the relative distance between the two lobes.

TRANSDUCERS: SENSOR UNITS FOR DETECTING CHANGES IN AIRSPEED AND ALTITUDE

A *transducer* is defined by Webster as "a device actuated by power from one system and supplying power in the same or any other form to a second system." In most missile applications, the transducer is used to change a mechanical action caused by a pressure change into a reference voltage proportional to the mechanical change. This action is accomplished by linking the mechanical action to a wiper arm on a linear potentiometer. In the missile field, a transducer is most generally used as an end instrument of the telemetering system. Transducers are the components used to translate changes in airspeed and altitude into electrical control signals.

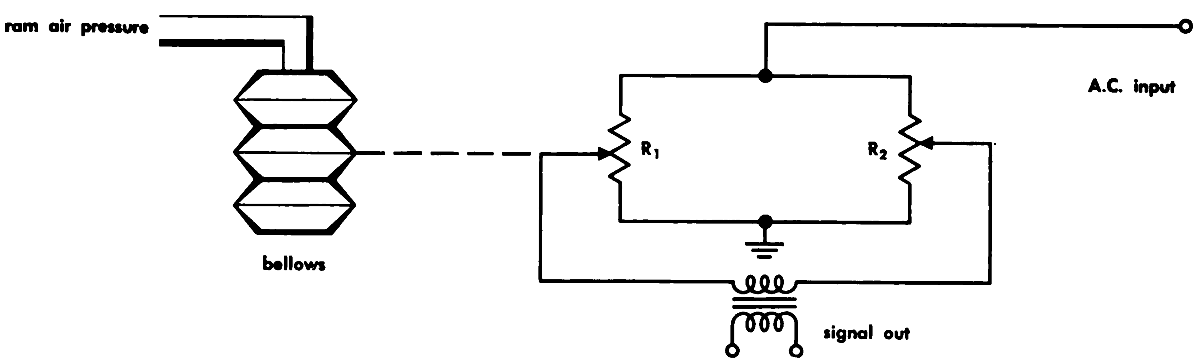
Air Speed Transducers

An airspeed circuit is used to maintain desired constant indicated air speed of a missile. Changes in air speed may be accomplished either by changing the throttle setting or by changing the pitch attitude of the missile. The air speed transducer may consist of bellows mechanically attached to the wiper on a linear potentiometer (pot) or a bellows mechanically attached to the rotor or stator of a selsyn. The bellows is actuated by ram air pressure. Therefore, when the air speed changes, the bellows changes (expands or contracts) causing a change in the setting of the pot wiper or selsyn rotor, or stator. The potentiometer type of air speed transducer is generally used in a bridge type circuit as illustrated below.

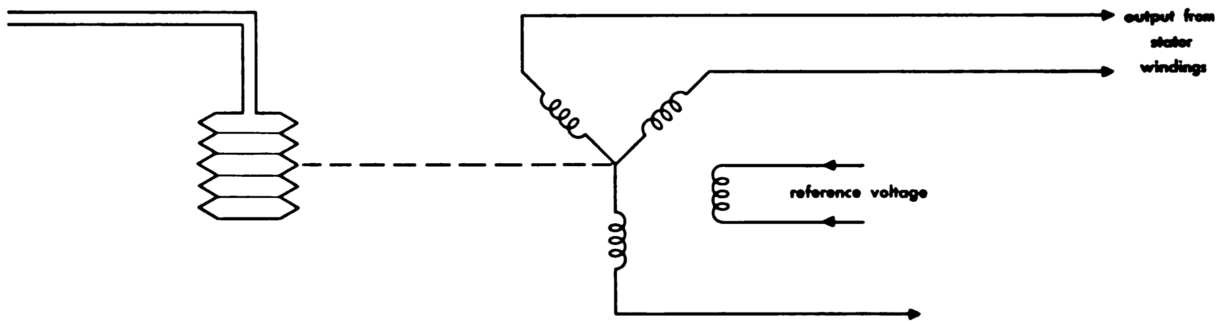
The wiper of potentiometer R_1 is attached to the bellows and is repositioned on the pot whenever air pressure changes occur. The wiper of potentiometer R_2 is adjusted at a point on the pot in such a way that the bridge will be balanced at the desired reference air speed.

The bridge balancing operation is performed before the missile is launched by applying air pressure equivalent to the air pressure of the reference air speed. The wiper on R_2 is fixed at this particular point by means of a set screw or similar device.

The next illustration shows a selsyn-type air speed transducer. The stator of the selsyn is mechanically connected in such a way that



Resistor-bridge-type airspeed transducer



Selsyn-type airspeed transducer

expansion or contraction of the bellows causes a rotation of the selsyn stator.

Once a sensor unit detects a change in the attitude of a missile, the newly acquired intelligence must be transmitted in a useful form. Pickoffs serve this purpose.

PICKOFFS: DEVICES THAT MAKE INTELLIGENCE OF SENSOR UNITS USEFUL

A pickoff is a device which produces a useful signal from the intelligence developed by a sensor. This signal must meet the particular requirements of the servo loop it is serving (as to phase, amplitude, loading effect, etc.).

The sensing devices pertinent to the attitude control of a missile generally make use of an angular or linear displacement proportional to some quality being measured or to the difference between the amount of the existing quality and some desired amount needed for operation. The physical dimensions of the displacement will vary; each variation requires different types of pickoffs to meet different needs in the servo loop.

The pickoff must first have an output sense. That is, it must be able to distinguish the direction of the displacement and produce a signal indicative of the direction. In electrical pickoffs this is frequently done by using a phase or polarity difference. The ideal pickoff should also have a maximum output for a minimum movement of the pickoff. It should have a linear output. It should also have a minimum torque or friction loss which would be reflected back to the sensor element.

Small size and lightweight construction are additional requirements imposed on pickoffs to be used in missiles. Pickoffs also should have a fine null range, and electrical pickoffs should have no phase shift through a varying displacement.

A number of pickoffs have been used or experimented with in missile control systems. The majority of these are electrical. Some pneumatic devices have been used in systems in which it was desired to keep electrical components to a minimum, but they have not proven too satisfactory and are not considered here.

Of the electrical pickoffs generally in use, the following are the most common:

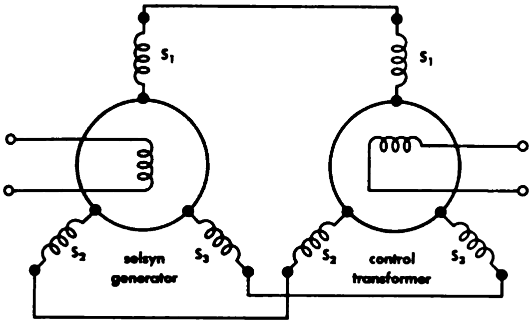
a. Selsyn.
b. Potentiometer.
c. Reluctance.
d. Capacitance.

Each of these is found with variations to make it suitable to specific applications in which its particular characteristics are needed.

Selsyn Pickoffs

A selsyn pickoff (also called synchro, autosyn, microsyn, etc.) usually consists of a pair of selsyns wired as a generator and a control transformer. They give an accurate electrical identification of angular movement, can be made small and light, and have a constant phase for a varying displacement; but they lack sensitivity for minute variations in displacement.

Two typical selsyn pickoff systems are shown in the accompanying illustrations. When a selsyn generator and a selsyn control transformer are properly connected, they form a complete pickoff system. As shown in the diagram at the right, a differential selsyn may be included between the generator and



Selsyn hookup

control transformer. The application of the differential selsyn is explained in detail later.

Both schematics represent the electrical zero position. For a control transformer, electrical zero is defined as that position of the rotor in which no voltage is induced in the rotor windings from the stator windings. This condition occurs when the axis of the rotor is perpendicular to the axis of the S_1 winding.

OPERATION OF A CONTROL TRANSFORMER. A control transformer and a selsyn generator are connected in the electrical zero position. Note the relative positions of the two rotors in the accompanying diagrams. No voltage is induced by the S_1 coil since the rotor and coil are perpendicular. The currents in the S_2 and S_3 coils are equal and opposite and therefore induce voltages in the rotor which are equal and opposite. The net effect is that no voltage is induced in the rotor in this position.

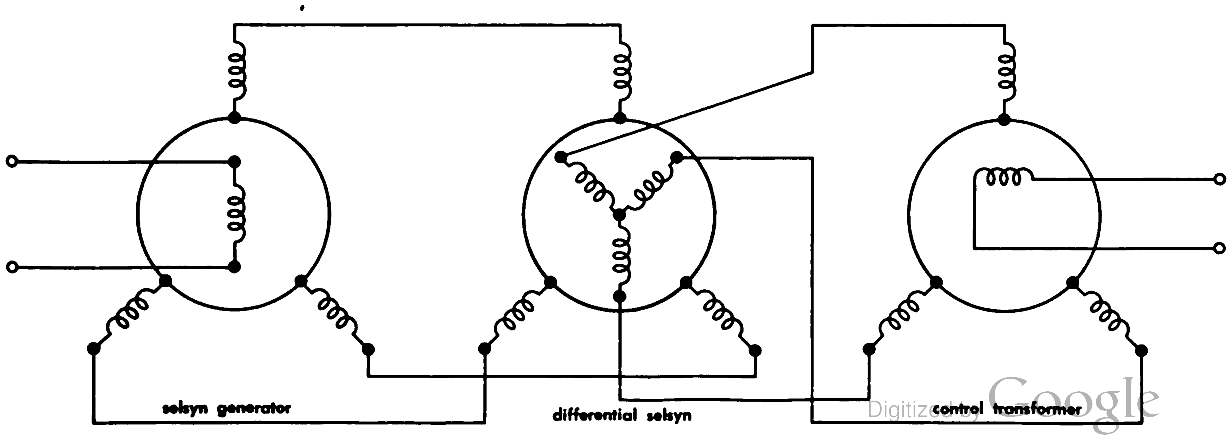
When the shaft of the control transformer is turned 90 degrees clockwise or counterclockwise, the position of S_1 and the rotor is such that maximum voltage is induced in the control transformer rotor. The currents in the S_2 and S_3 coils are equal and in the same direction, and they therefore induce voltages in the rotor which are equal and of the same polarity. The net result is that maximum voltage is induced in the rotor in these positions and the induced voltage is either in phase or 180 degrees out of phase with the voltage of the rotor of the selsyn generator.

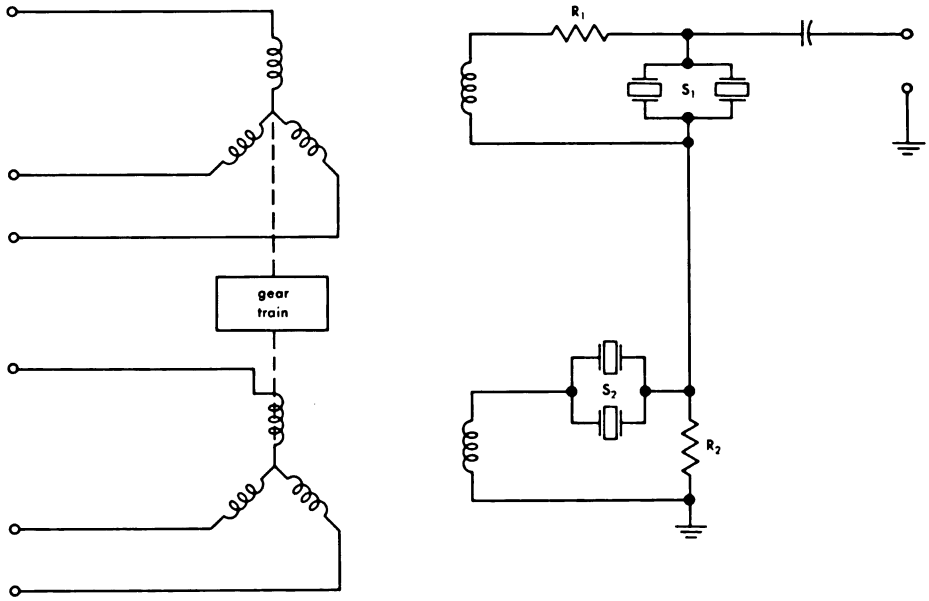
The output voltage obtained from a control transformer is called an error signal, because the magnitude and phase of this voltage is an indication of how much and in what direction the two rotors are out of correspondence. Turning the rotor of the control transformer so that its position is perpendicular with that of the generator rotor cancels or nulls the error signal. This position is sometimes referred to as the null position. The ordinary selsyn pickoff system has a wide null position; that is, it lacks sensitivity for very small displacements. For this reason two-speed selsyn systems are quite often employed in order to increase their pickoff sensitivity.

TWO-SPEED SELSYNS. In a two-speed selsyn system, the *fast selsyn* of the pair is mechanically attached to the *slow selsyn* through a step-down gear train as shown on page 192. Thus, with an appropriate relaying arrangement, the error signal can be picked off the fast speed selsyn during small displacements of the rotor shafts. When the error signal exceeds a certain preselected value, the relay positions change, and the error voltage then is taken from the slow selsyn rotor. Sensitivity is thus increased since the fast selsyn measures an appreciable angular displacement for small displacements of the slow selsyn.

Another type of two-speed selsyn system does not require any switching or relaying

Selsyn hookup

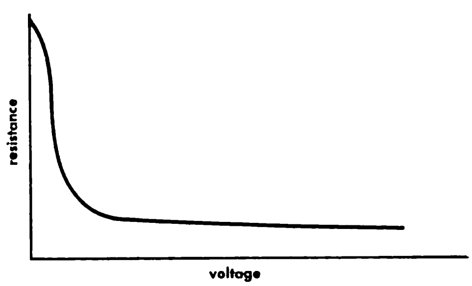




Two-speed selsyn system

between selsyns. In this system the characteristic curve of a selenium rectifier for forward resistance versus voltage applied is utilized. As shown on the graph below, the resistance of the rectifier is high for small voltages and decreases as voltage is increased.

When the angular displacement of the slow selsyn rotor is large with respect to the null position, the slow selsyn's error voltage is large. This voltage is divided between the selenium rectifier and the resistor. Since this voltage is large, the resistance of the selenium rectifier is small in comparison to that of the resistor. Therefore, most of the voltage drop is across the resistor. The fast selsyn output can be large or small at this time since it cycles several times for each revolution of



Resistance vs voltage curve of selenium rectifier

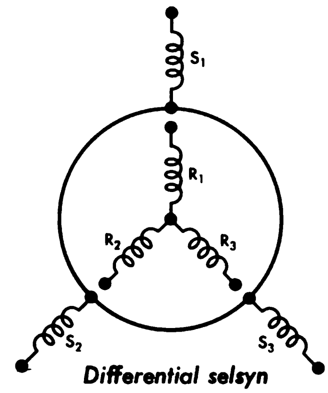
the slow selsyn. The fast selsyn output voltage is limited by selenium rectifier S_1 and resistor R_1 to two or three volts; however, the output voltage to the succeeding circuit is the sum of the voltage across R_2 and S_1. As long as the slow selsyn error output voltage is greater than two or three volts, it predominates over the total servo output voltage. As the system comes closer to its null position, the majority of the slow selsyn output is dropped across S_2.

When the system is three or four degrees from null position, the slow selsyn voltage across R_2 is so small that it contributes little to the error signal. Therefore, the error signal is essentially that of the fast selsyn alone.

DIFFERENTIAL SELSYNS. A differential selsyn performs the same function electrically that a mechanical differential performs in a mechanical system. The differential in a mechanical system connects three shafts so that one shaft turns an amount which is equal to the difference between the amounts that the other two turn. A selsyn differential generator subtracts two inputs, and a selsyn control transformer indicates the difference between the two inputs. The action of the differential selsyn can be reversed so that two inputs are added by a control transformer.

The stator of a differential selsyn is similar to that of an ordinary selsyn. It consists of three sets of Y-connected coils wound in slots. The slots are spaced 120 degrees apart around the inside of the field structure. The rotor of the differential, on the other hand, is completely different from that of the ordinary selsyn as the accompanying illustration shows. It is cylindrical in shape and has three sets of Y-connected coils wound in slots which are spaced 120 degrees apart around the circumference. Three slip rings connect the rotor leads to the external circuit as in the ordinary selsyn. In the electrical zero position, the rotor coils R_1, R_2, and R_3 are turned so that they are aligned with the stator coils S_1, S_2, and S_3, respectively.

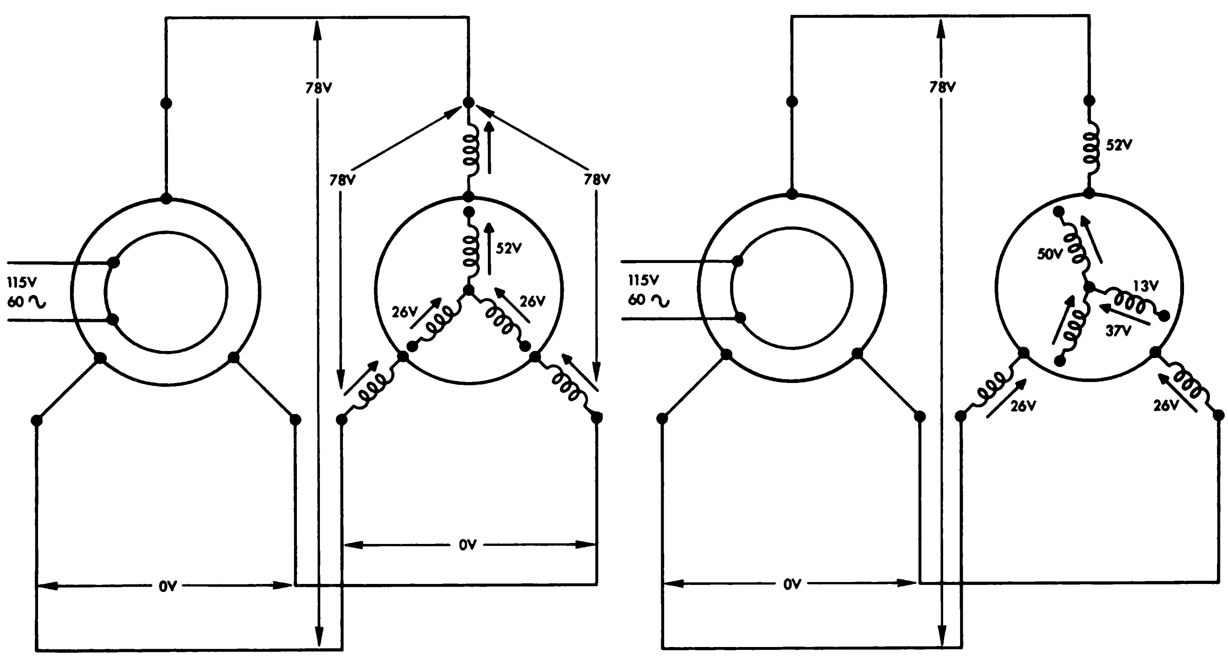
Like the ordinary selsyns, the differential operates on the transformer principle. The stator acts as the primary; the rotor, as the secondary of a 1:1 transformer. Because of the air gap, more turns are wound on the rotor coils than on the stator coils in order to achieve the 1:1 ratio. Therefore, the stator must always be used as the primary and the rotor as the secondary of a differential selsyn.



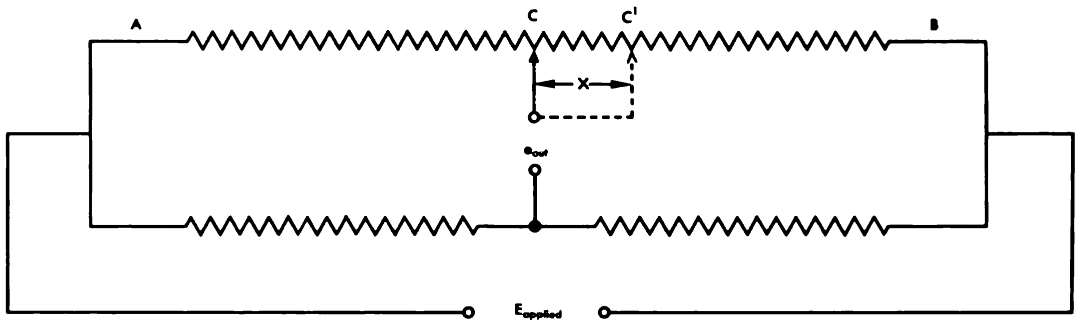
Differential selsyn

The transformer action of a differential selsyn is illustrated in the diagrams below. Assume that both rotors are in the electrical zero position, the stators are connected as shown, and the rotor of the differential selsyn is open. Since the stators are connected together and are in parallel, the voltages in both sets of stators are equal to each other and are in phase. The stator voltages of the differential, through transformer action, cause similar voltages to be induced in each of the rotor windings.

If the rotor of the differential were turned to the 15-degree position and the rotor of the



Transformer action in a differential



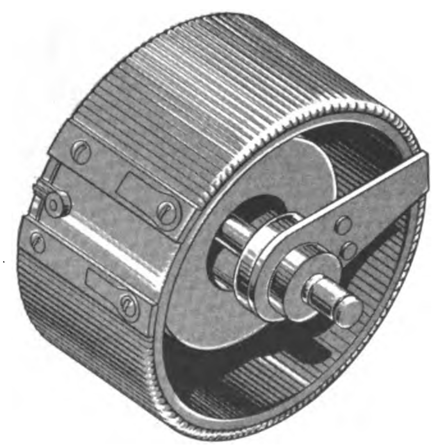
Basic bridge network using a potentiometer

selsyn generator remained fixed, the stator voltages would be equal as before. However, the voltages induced in the rotor coils of the differential are less because all the flux from each stator winding no longer cuts its corresponding rotor winding.

A differential selsyn receives two inputs, one is electrical and the other is mechanical. It subtracts one input from the other and transmits the difference.

Potentiometer Pickoffs

A potentiometer is a tapped resistor in which the position of the tap can be altered by some sort of mechanical control. The standard high resistance potentiometer used in communications circuits uses a resistance element consisting of a thin film of carbon deposited on some insulating material. However, this element is not practical for servo use because the resistance of such a carbon



Wire-wound potentiometer

film changes with wear and with variation in temperature and humidity. Since a more rugged type is required, the potentiometers used in servos generally use resistance elements consisting of a number of turns of resistance wire.

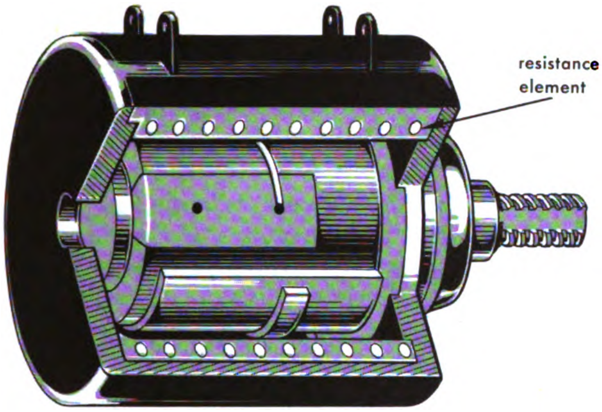
A potentiometer pickoff is easy to fabricate and install. It can be mounted in any position and works nicely in bridge-type control networks; however, it usually has a large loading effect upon the network. It also lacks sensitivity for minute variations in displacement which limits its application to certain sensors.

The degree of complexity of the bridge network depends upon factors other than the pickoff, but all are variations of the same basic bridge circuit shown above.

An output voltage is not developed when the potentiometer (AB) is arranged in a bridge network, and the wiper is at position C with the resistance equally divided on each side of the wiper. An output voltage is developed when the wiper is displaced some distance (X) to position C′.

Since the potentiometer is attached to the sensor so that its displacement is the distance "X," the output voltage varies directly with the displacement.

CONVENTIONAL POTENTIOMETERS. Most potentiometers used in servos are essentially the same, consisting of wire resistors and movable sliding contacts. In the conventional type, shown in the illustration, Wire-wound Potentiometer, the resistance wire is wound on an insulating strip. The strip, together with the wire, is called a resistance card. The card is bent so that it forms an arc of



Helipot, showing resistance element wound in the helix

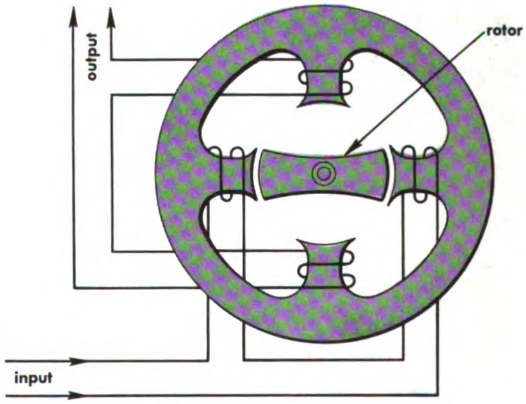
nearly 360° and is mounted so that the slider makes contact with the winding along one edge of the strip. The slider usually can rotate continuously, but it loses contact with the resistor over a small arc.

In servo usage, a potentiometer should be thought of as a variable voltage source rather than as a variable resistor since an input voltage is applied to the resistance element and a fraction of this voltage appears as the output between the slider and one end of the resistance. Unfortunately, the output of a potentiometer does not change smoothly as the slider is moved. Instead, the output voltage changes in jumps, each jump being equal to the voltage difference existing between adjacent turns of wire. A potentiometer having 1000 turns of wire on the resistance element is said to have a resolution of 1 part in 1000 or a resolution of 0.1 per cent. A resolution of 0.1 per cent means that the smallest change in the output voltage is 1/1000 of the input voltage.

In order to improve the resolution, the resistance element is sometimes wound in a helix as shown in the helipot illustrated above. The slide may make as many as ten turns in covering the whole element, and although this construction allows many more turns on the resistance element, it suffers from the disadvantage that the slider cannot rotate continuously. Since the change in voltage along the resistance element is proportional to the change in resistance, it is simple to make a nonlinear potentiometer by using a nonuniform resistance card.

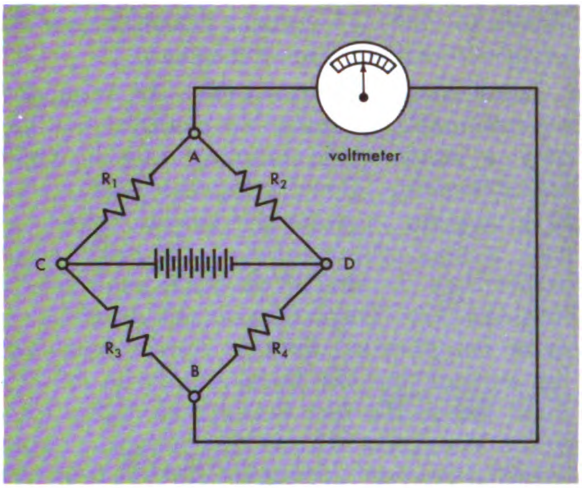
INDUCTION POTENTIOMETER. An induction potentiometer is similar to a selsyn but has only a single rotor winding and a single stator winding. By shaping the pole pieces. it is possible to make the output voltage of such a unit proportional to the shaft angle rather than the sine of the shaft angle over a limited range. The induction potentiometer is more complicated than a conventional potentiometer and can be used only on AC. Its merit lies in the fact that it has higher resolution than most wire wound potentiometers.

MICROSYN. A microsyn, as shown in the illustration below, is used occasionally as an induction potentiometer. Its chief merit is that there are no electrical connections to the rotor; consequently, there are no brushes. The magnetic field of the input windings magnetizes the iron rotor. The magnetic field of the iron rotor then induces voltage in the output windings unless the rotor is in its zero position. Over a limited range, the output voltage is proportional to the displacement of the rotor. The microsyn, like other magnetic elements, requires alternating current.



A microsyn

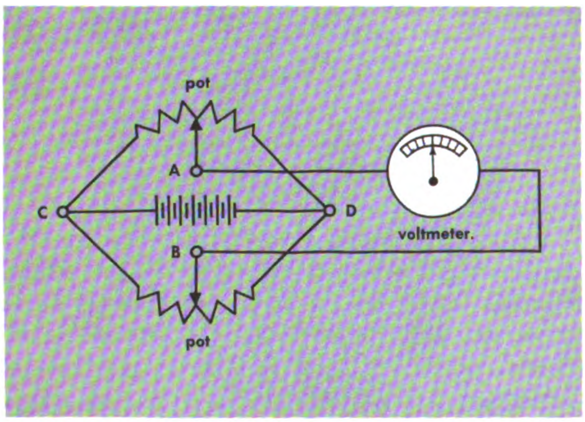
POTENTIOMETERS IN BRIDGE CIRCUITS. Potentiometers are sometimes used in bridge circuits to derive a signal proportional to direction and magnitude of deviation in all three control axes. For a brief analysis of resistance bridges, let's refer to the bridge in the following illustration. Equal resistances R_1 and R_2 are connected in series across a DC



Simple DC resistance bridge circuit

voltage source. Also connected in series across this voltage source are equal resistances R_3 and R_4. R_3 and R_4 are therefore in parallel with R_1 and R_2. Since R_1 and R_2 are of equal value, applied voltage is divided equally across them. In a similar manner, the voltage drop across R_3 is equal to that across R_4. In this circuit then there is no voltage difference between points A and B.

The bridge circuit is essentially the same if the resistances are replaced by two potentiometers, one in each branch of the circuit as shown in the next illustration. When the potentiometer wipers are placed on the exact electrical center of the resistances, the voltmeter shows no voltage difference between points A and B. When the wiper is moved to the left, the bridge becomes unbalanced and

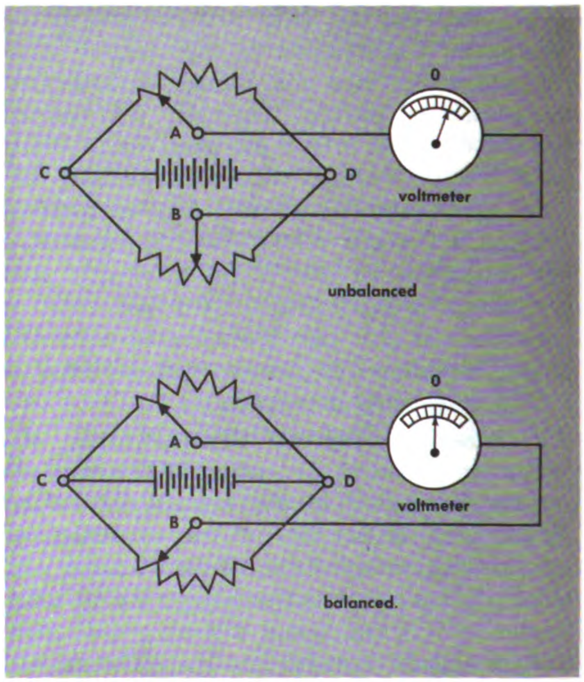


DC bridge circuits with pots

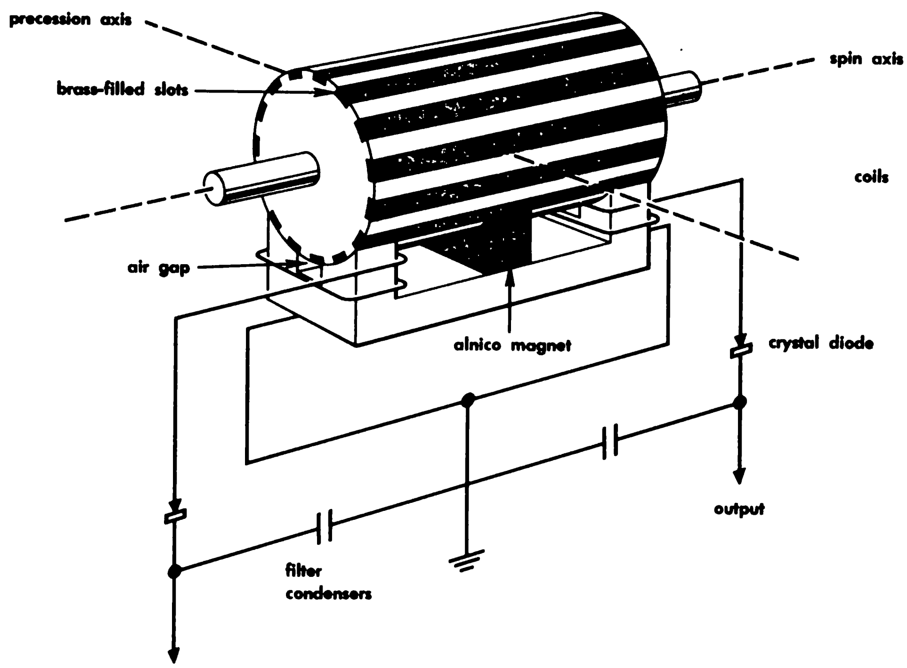
voltage difference exists between wiper A and wiper "B." The amount of voltage recorded is in proportion to the distance which the wiper has moved across the winding, provided the resistance of the winding is distributed uniformly from one end to the other. When a balanced bridge network is used for control purposes, the voltage which is developed at the wiper when the bridge is unbalanced is called a control signal.

If the potentiometer wiper "B" is moved across the winding in the proper direction and the proper amount, the bridge circuit can be rebalanced without returning wiper "A" to its original position. This fact is demonstrated in the illustration below. Potentiometer "B" has, in effect, produced a control signal in the opposite direction, and this signal has balanced out the signal supplied by potentiometer "A." The polarity shown by the voltmeter when the bridge is unbalanced depends on the direction of movement of potentiometer wiper "A."

The simple bridge circuit described in the previous paragraphs is energized by direct current; therefore, the direction of the control



Unbalanced bridge supplying a control signal and a rebalanced bridge



Internally operated reluctance pickoff

signal is indicated by the polarity at the voltmeter. In actual practice, however, it is more practical to energize the bridge with alternating current. The direction of the control signal in an AC bridge circuit must therefore be determined by the phase of the alternating current.

Variable Reluctance Pickoffs

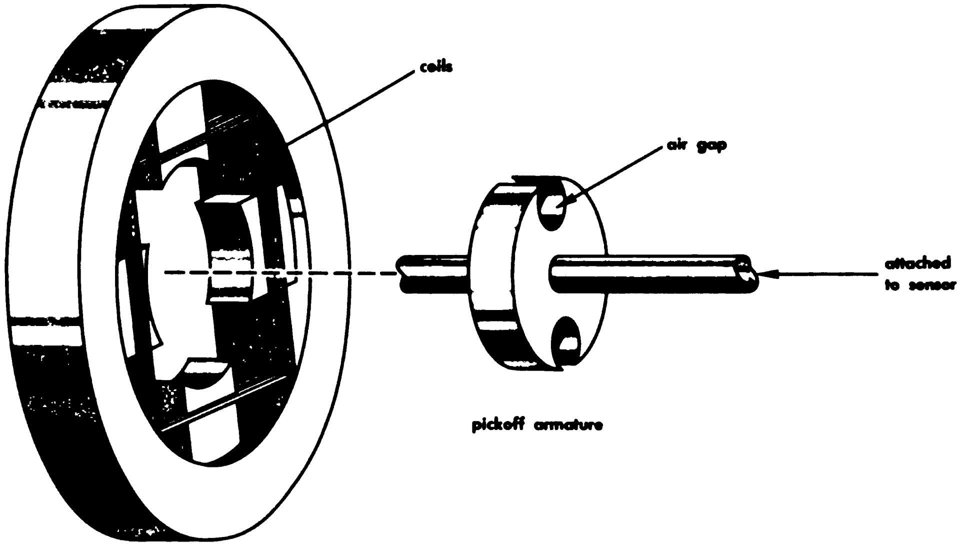
Variable reluctance pickoffs have been extensively investigated. Such pickoffs have been used with sensors which have small displacement values or oscillatory movements. They give the largest output per unit displacement and with a minimum loading effect, but they also require stable oscillators, thus making space and weight demands on the missile.

A number of types of reluctance pickoffs have been developed. Since reluctance pickoffs require an oscillating voltage for their operation, they can be classed by the method used to supply this voltage. An internally oscillated pickoff is illustrated above. The figure shows a rate gyro rotor and its reluctance pickoff. The pickoff consists of an E-shaped metal mass with coils wound around its extremities and a permanent magnet located in the center. The gyro rotor is a ferrous material which has been slotted and the slots refilled with brass to restore the lost weight.

The magnetic force set up by the magnet causes a flux to flow through each end of the pickoff mass, through the gyro rotor, and back into the magnet. The gyro rotation causes regular variations in the reluctance of the flux paths as the brass or ferrous metals pass over the end pieces. This causes a regular variation in flux density to be established, inducing an AC voltage in the coils. However, the voltages in the coils cancel each other out.

When the gyro precesses, the air gaps at each end vary oppositely in proportion to the acceleration and cause different voltages to be induced in the coils. The difference in these two voltages is then proportional to acceleration, and after being rectified the difference is the signal voltage. The output "sense" shows up here as a difference in polarity.

Another type of reluctance pickoff, shown next, consists of a stator containing two pairs of coils mounted in space quadrature. One pair of coils is supplied with a constant



Externally operated reluctance pickoff

amplitude AC voltage from a reference oscillator. Voltage is induced in the second pair of coils through a split iron core rotor coupled to the gyroscope gimbal. As the angle between the rotor and the pickoff coils changes due to variations in missile attitude, the amplitude of the AC voltage induced in the pickoff changes proportionally; the phase of the voltage induced differs by 180 degrees, dependent upon the direction of the missile attitude change from the zero position.

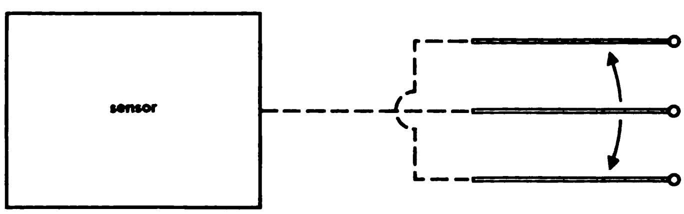
Capacity Pickoffs

We'll conclude this discussion of pickoffs with a brief mention of *capacity pickoffs*. A variable capacity pickoff makes use of a movable capacitor plate located between two fixed plates. Upon movement of the center plate by the sensor unit, the capacities of the two condensers are varied and their output, when wired into the appropriate circuit, is useful. This pickoff is extremely sensitive.

SENSOR UNITS: THE SECOND REQUIREMENT OF MISSILE CONTROL SYSTEMS

In this chapter, you became acquainted with the second requirement of a missile control system — sensor units. You found that gyroscopes, altimeters, and transducers provide a means of detecting any departure of a missile from the desired reference.

The next chapter is concerned with computer units which help to satisfy the third requirement of an automatic control system. Computers, along with other types of units, convert error information into a form that can be used to regulate the controlling devices of a missile.



Capacity pickoff

Computer Units of Control Systems

Up to this point we have considered methods of detecting various missile deviations from desired conditions. In the general control block diagram early in the chapter, we found that these deviations are detected by means of the *sensor* block operating with the *reference* block. We know that the output of a sensor is referred to as an *error*, which of course represents an existing missile deviation from desired conditions. The most important deviations are changes in a missile's attitude. Other conditions — such as changes in missile airspeed, altitude, angular acceleration, or forward and sideward acceleration — also can be detected. Deviations from a desired condition may be caused by changing winds, changing air pressure, and/or changing engine thrust.

Although the output of a sensor represents a situation which must be corrected, this output seldom goes directly to the actuator which effects a missile change. The error signal must first be "operated" on and amplified. This operation consists of changing the signal to represent additional information which is required for proper control functioning. The operation is represented by the "computer" block which follows the sensor block in the general block diagram.

A computer in a control system is comprised of three general types of components, which are:

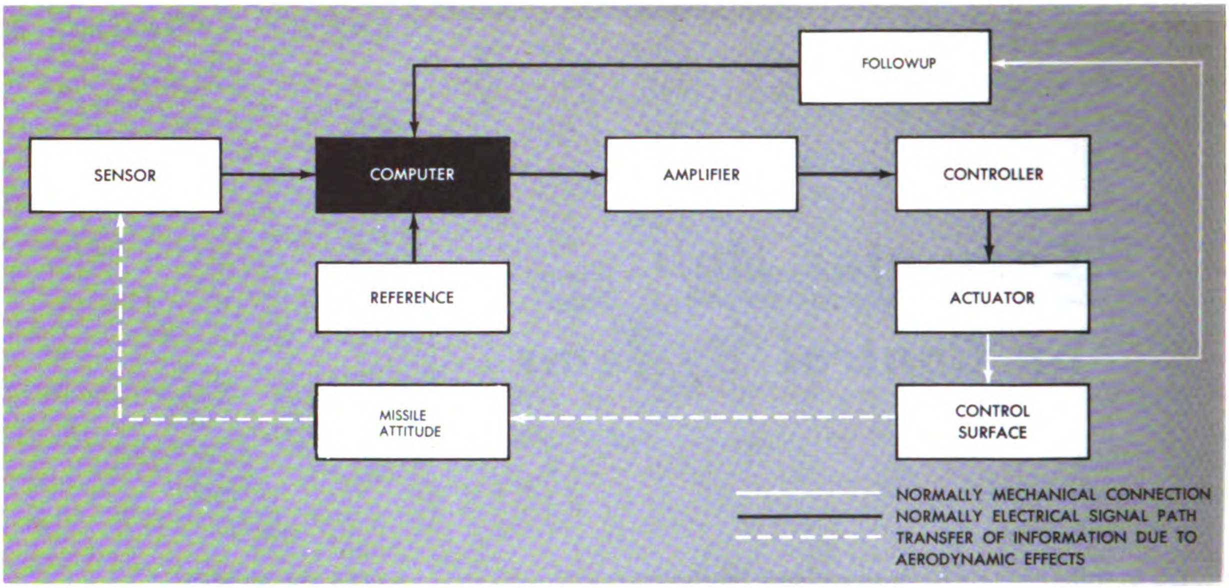
1. Mixers.
2. Integrators.
3. Rate components.

TYPES OF MIXERS

A mixer is any circuit or device which combines information from two or more sources. Every missile control system, which operates as shown in the basic block diagram on the following page, must combine several signals. Reference to the diagrams shows that a followup signal is combined with the sensor error signal. A rate or integral signal, which is produced within the computer, is often combined with the sensor error signal. In some missiles several inputs are mixed together. This mixing occurs when the required aerodynamic reaction depends on more than one variable. For example, a system may be controlling the pitch of a missile; this pitch could be influenced by missile attitude, altitude, and airspeed. A separate sensor and reference would be used to detect each of the three variables. All of these signals would be mixed together so that each variable would have the required influence on the pitch actuator.

Requirements of Mixers

For a mixer to function properly, it must combine signals in the correct *proportion*, *sense*, and *amplitude*. If a followup signal is combined with an attitude signal, the followup signal must be of a certain strength as compared to the error signal. This proportion can be indicated by means of *weighting factors*. Such a means determines the comparative *weight* or *influence* of a signal. Suppose that three signals — X, Y, and Z are to be summed



Basic missile control block diagram

and that only one-third of Y and one-half of Z should be used as compared to the strength of signal X. One-third of Y and one-half of Z are weighting factors.

A sensor signal must produce action in the desired direction. The sense of a followup signal must be such as to have a counteracting effect on the output. Also, the total output must be of proper amplitude for the right amount of control. Normally the signals either add or subtract, although other functions can be performed. These processes which are necessary for proper mixing can be considered a computer function.

The type of mixer used depends mostly on the type of control system. For example, most systems are basically electronic; therefore, most mixers are electronic. Mixers also may be mechanical, pneumatic, or hydraulic. These types are discussed on the following pages, for a knowledge of different types of mixers will help you to recognize a mixer function in missile circuitry. Such knowledge also will help you to analyze mixer malfunctions which may evolve.

Electronic Mixers

Fundamentally, the principles of any electrical mixer reduce to basic network and vacuum tube theory. Mixers consist of either an impedance network of resistors, inductors, and capacitors, or of several vacuum tube stages. The information to be combined is represented by the amplitude and phase of each voltage.

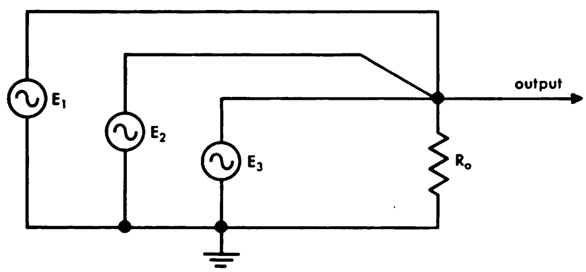
Mixers combine voltages from such sources as the output from pickoffs, rate components, integrators, followup generators, or guidance circuits. Whatever the sources, they are indicated in the following diagrams by the standard symbol of an AC generator.

Electric signals in a control system are usually AC. These signals are then changed to DC before they are applied to the actuators which move the control devices.

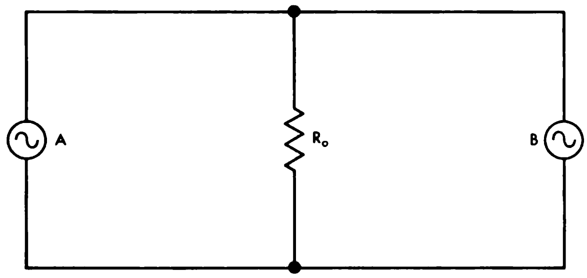
The phase of the AC signals determines the sense. If a certain two signals must make the same effect on a rudder, they must be in phase in the mixer. Occasionally, capacitors or inductors are necessary to alter the phase of signals so that they will mix in exactly the correct phase. If DC signals are used, they must be mixed with the proper polarity.

PARALLEL MIXING. The most simple method to combine voltages is to apply them across a common impedance by means of a junction. The voltage sources are effectively in parallel as shown in the next figure.

This circuit is normally altered in actual practice. The change consists of an added

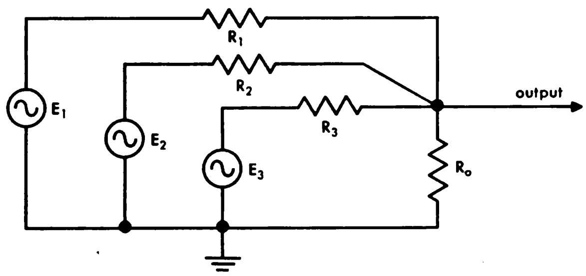


Basic parallel-type signal mixing circuit



Simple network with two voltage sources

impedance in each branch as shown in the diagram, Actual Parallel-type Signal Mixing Circuit. These added resistors reduce the voltages applied from each branch to the common impedance "R_o." In analyzing a circuit, such resistors usually are called **attenuators.**



Actual parallel-type signal mixing circuit

Actually, there are three possible reasons for these resistors. First, they reduce the current flow from the sources when an increase in the resistance of "R_o" is not feasible. This reduction of load current decreases the power consumption and sometimes makes the output of the pickoffs (voltage sources) more linear.

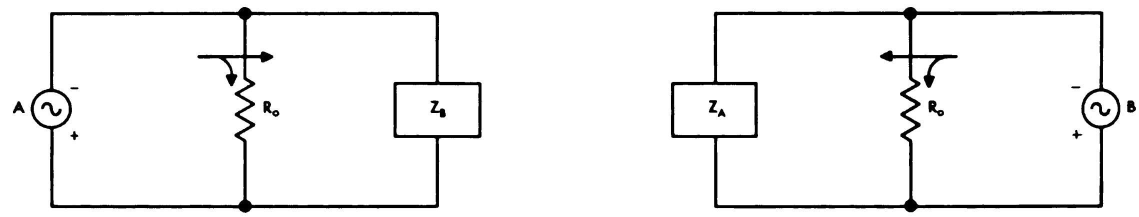
Second, these resistors allow the designer to adjust the amount of signal which should be applied to "R_o." Such adjusting makes possible the mixing of the signals in the proper ratio (or, with the proper weighting factors), a condition which is important. Often, it is easier to alter a pickoff output by using a series resistor of a certain value than to acquire a pickoff with a certain output.

Third, these added resistors reduce errors caused by coupling between the signal sources. This coupling is produced by current from one source changing the voltages which would normally be produced by the other source. Such an error occurs from a source in which the internal impedance varies, such as a potentiometer pickoff or an inductance pickoff in which the inductance field builds up with current flow. The cause of this error can best be understood after a review of the superposition theorem.

The operation of a circuit with many voltage sources generally is much harder to visualize than the operation of one with just a single generator. One of the best ways to visualize the former type is by the use of a standard electrical law of operation, the superposition theorem: If a network of linear impedances is energized by two (or more) generators, the current or voltage at any specified point in the network is expressed as the sum of the voltages or currents that each generator would produce if it alone were connected to the network and if the other generator were replaced by its internal impedance. This theorem can be applied in analyzing the circuit in the figure above.

The next circuits are a breakdown of the above circuit. The breakdown will help us in considering currents and voltages from each source separately. The electron flow indicated by the arrows results from assuming an instant when the polarity of the AC generator is as shown. In each case the flow represents current due to a single voltage. When the circuit is considered with both voltage sources, the actual current in each path is the vector sum of the components that would result from the sources considered separately. The *vector sum* is the result produced when voltages combine in a manner that depends upon the amplitudes and phase differences.

The application of the superposition theorem proves that the above circuit can be used as a mixer and that the series resistors can be used to adjust the ratio of the signals. The total voltage across "R_o" depends on

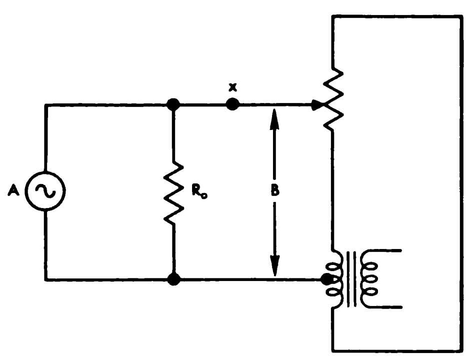


Circuit breakdown, considering each applied voltage separately

the total current. Since the total current is the vector sum of the branch currents, the output voltage depends not only on the strength of the voltage sources but also on the impedance in each branch.

As mentioned before, changes in internal impedance of pickoffs which are interconnected may result in signals of incorrect amplitude. Suppose a potentiometer pickoff signal is combined with a signal from source "A" as shown in the illustration below. The current developed from source "A" depends on the total impedance in the circuit. When this impedance is varied by movement of the potentiometer, the current from source A varies. The varying of the current changes the source "A" voltage applied to "R_o," even though the cause of the change does *not* represent a *change of the information* represented by voltage "A."

If a resistor were inserted at point "X," the current variation due to movement of the potentiometer would be reduced, thus there would be less effect on the signal voltage at source "A."



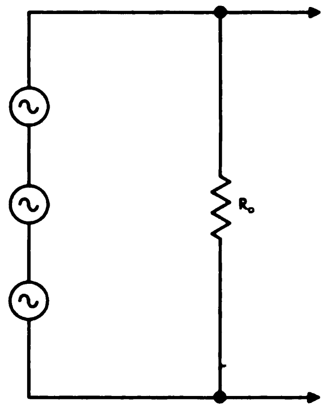
Circuit with source in which internal impedance will vary

SERIES MIXING. Voltages also are combined by applying them in series with a common load, as shown in the diagram, Basic Series-type Signal Mixer Circuit. A disadvantage of this method is that all the load current must flow through each source. This fundamental circuit is normally altered as shown on the extreme right. The resistors are included for the same three reasons as in the case of parallel-type mixers.

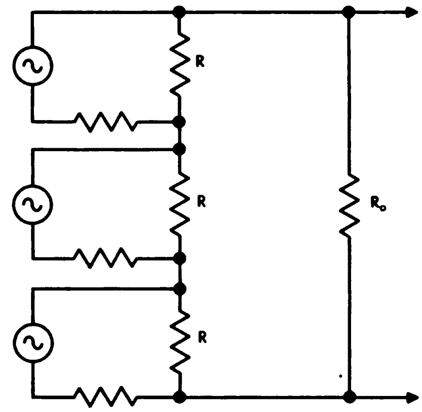
The sum of all voltage drops and rises around a closed loop must equal zero. Therefore, in the basic circuit, the voltage across "R_o" is the sum of the generator terminal voltages. In the actual circuit, voltage across "R_o" is the sum of the voltages across the resistors labeled "R."

RESISTANCE BRIDGE MIXER. A resistance bridge network mixes position information from two sources. Suppose that position information of a gyro and a control surface must be detected and combined. Two sliding contacts are used in the bridge, as shown on the right. One contact is mechanically connected to the gyro gimbal and one to the control surface. Physically, the potentiometers are located quite a distance apart. The bridge serves the purpose of two electrical pickoffs and a mixer.

The purpose of the circuit is to mix a voltage proportional to gyro displacement with a voltage proportional to control surface position. These two particular signals are often combined in systems since one signal represents the input error signal and the other is the followup signal. The potentiometers are adjusted in such a manner that if the missile is in the proper attitude and the control surface is streamlined, the bridge is balanced; therefore, no signal is sent to the amplifier to cause a



Basic series-type signal mixer circuit



Actual series-type signal mixer circuit

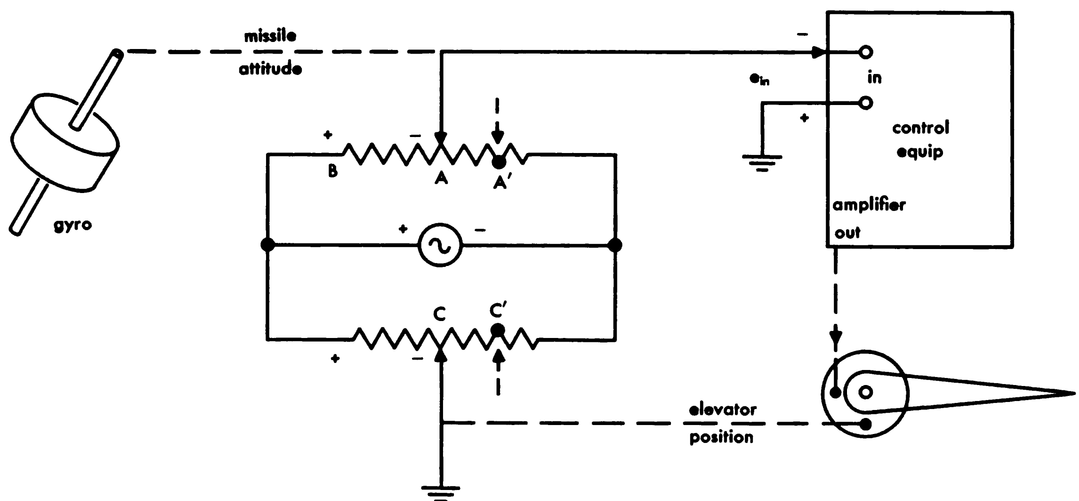
correction. Assuming an instant when the polarities of the AC voltages are as shown, the voltage across "AB" is exactly cancelled by "BC," so "e_{in}" is zero.

If the missile deviates, a voltage proportional to the distance the upper slider moves (AA′) is applied to the amplifier (assuming a linear pot and no load current flowing to the amplifier). In this assumed instant, the voltage is negative. If the missile is in the correct position but the control surface is not yet streamlined, a voltage proportional to the distance the lower slider moves (CC′) is applied to the amplifier. It is positive *at the assumed instant.* Normally, both actions occur at the same time, thus the voltage applied to the amplifier is the resultant of the two voltages. In this application, the mixer performs a subtraction since the signals caused by the displacements tend to cancel out. The resultant "e_{in}" is negative since the deflection of the upper pot is shown greater than that of the lower pot. (The control surface would then be deviating further from streamline.) That the resultant would be negative can be shown by the voltage equation for a closed loop. Starting at the grounded input terminal,

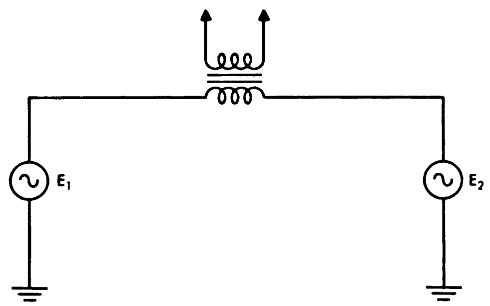
$$e_{in} - e_{ab} + e_{bc} = 0.$$

The weighing factors depend upon the mechanical linkages and on any resistors which are added to the bridge circuit.

Impedance networks often are used to divide a signal among more than one load. The same problems and network analysis applies to the



Mixing two signals by means of a resistance bridge

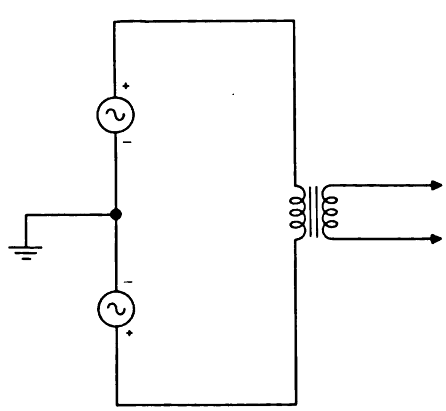


Series mixer using transformer output

separation as to the combination of signals. The problem becomes more involved when these loads are fed by other signals since a network must be designed to sufficiently "decouple" the unwanted voltages from each load.

Transformer mixing. The primary of a transformer can replace the output resistor "R_o" of series and parallel mixers. In such mixers, the secondary output is proportional to the vector sum of the input voltages.

Consider a transformer to which two voltages in series are applied as shown above. All resistors have been omitted. Assume that a phase check with an oscilloscope indicated that "E_1" was in phase with "E_2" with respect to ground. The question arises, Would the signal voltages aid or oppose in the transformer? The signals will oppose in the transformer since the two current components flow in opposite directions. Or, if you prefer, the lesser voltage will "buck out" part of the greater voltage. The answer can easily be seen by rearranging the figure and again assuming

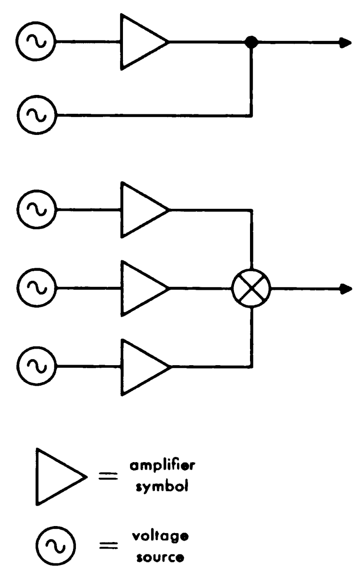


Series mixer with opposing inputs

an instant of time when certain polarities exist. This situation is shown in the figure on at lower left.

A transformer also can be used as a mixer by applying one voltage to the primary and one to the secondary. Any very low frequency or DC signals are applied to the secondary, because any AC signal applied to the secondary is induced back to the primary.

Vacuum tube mixers. Another method of eliminating coupling between voltage sources is by means of vacuum tube circuits, shown in the following diagram. A signal cannot feed back through an amplifier at the low frequencies used in control systems. Of course, the amplifier also performs the useful purpose of increasing the signal strength.

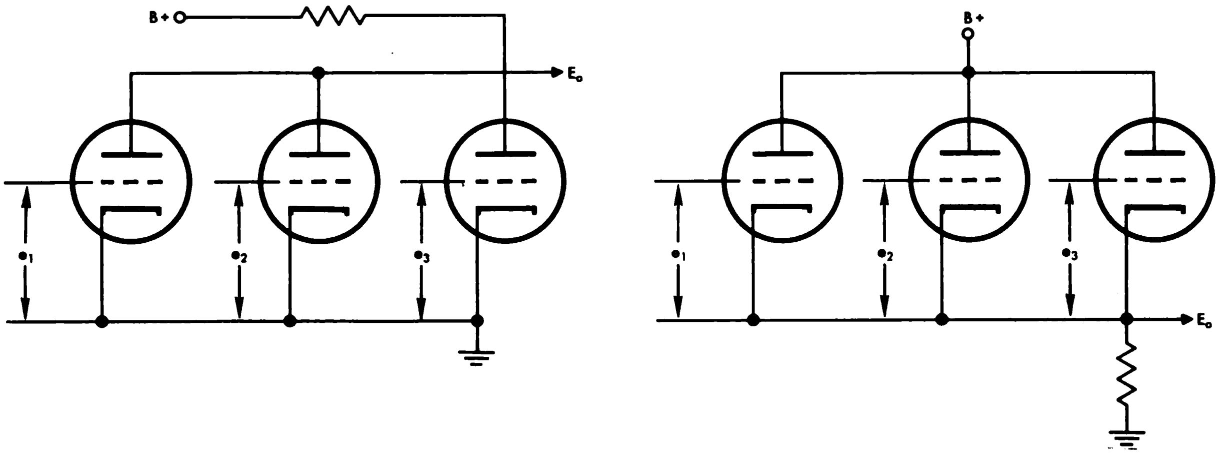


Parallel-type mixers using tubes for amplification and isolation

Two alternative circuits for the parallel type tube mixer are shown to the right above.

The number of possible inputs is not limited. Assuming the tubes are identical, similar inputs would have the same influence on the output. Weighting factors could be introduced by the use of pots at the input or resistors in the separate plate circuits.

The parallel amplifier circuit can be altered to produce a sum and difference output. Such a differential amplifier is shown in the diagram

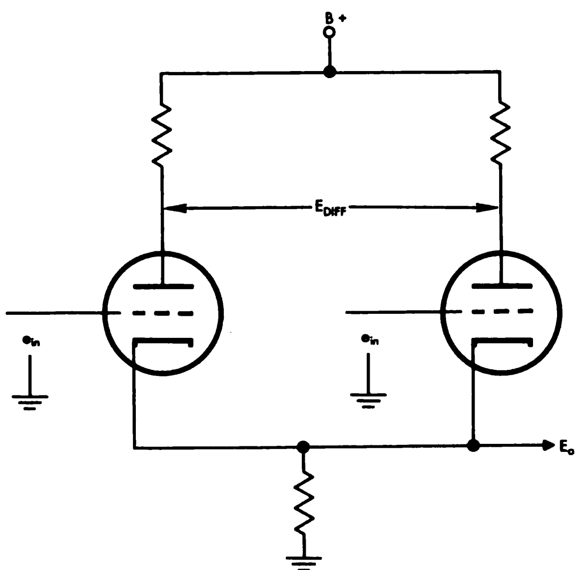


Parallel-type vacuum-tube adders

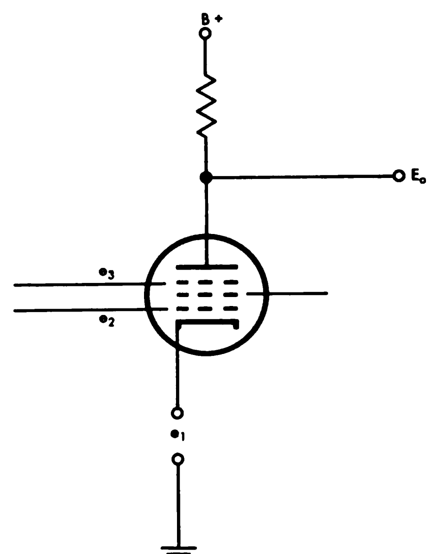
below. "E_o" is proportional to the vector sum of the input voltages, and the voltage between the plates is proportional to the difference of the input voltages.

Cathode followers also can be used in the same manner as amplifiers to prevent feedback or reflection of the output voltage. Cathode followers have the disadvantage of not amplifying voltage; however, they possess the advantage of very high input impedance which acts almost like an open circuit. When the cathode follower is connected to a pickoff (or any circuit), it presents practically no load.

An effective method of vacuum tube mixing is to combine signals within the tube itself. Several signal voltages can be impressed upon different electrodes of the tube to vary the plate current in the manner illustrated below. The effect of plate current variation due to one source has little effect on the other voltage sources since the impedances between the electrodes of any multigrid tube are high. Normally, such mixing is done by using different frequencies and is called "modulating." Mixing is common in communication-type electronic equipment such as radio, tele-



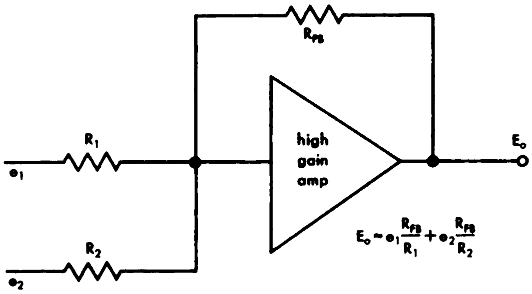
Differential amplifier



Multigrid mixer

vision, and missile guidance transmitters. Such mixing also uses the low frequencies of control systems.

The number of inputs is limited by the number of electrodes in the tube. The weighting factors depend on the characteristics of the tube and on the elements to which the input voltages are applied. Precautions have to be taken to avoid errors due to nonlinearity of the tube characteristics.



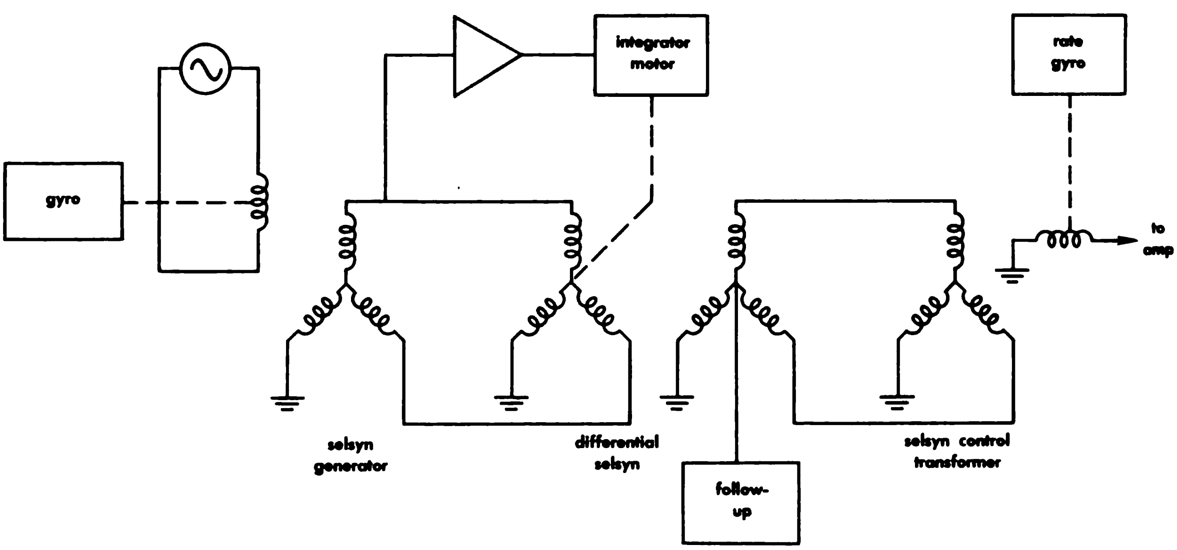
Parallel-type mixer using feedback amplifier

The parallel-type resistance mixer previously mentioned is often fed into a high gain amplifier which is made stable by the use of negative feedback. The feedback amplifier maintains almost a constant ratio between input and output because gain variations are reduced by feedback. The relationship depends almost entirely on the values of the resistors as shown on the left. The output depends entirely on the ratio of resistance, if the amplifier is assumed to have infinite gain.

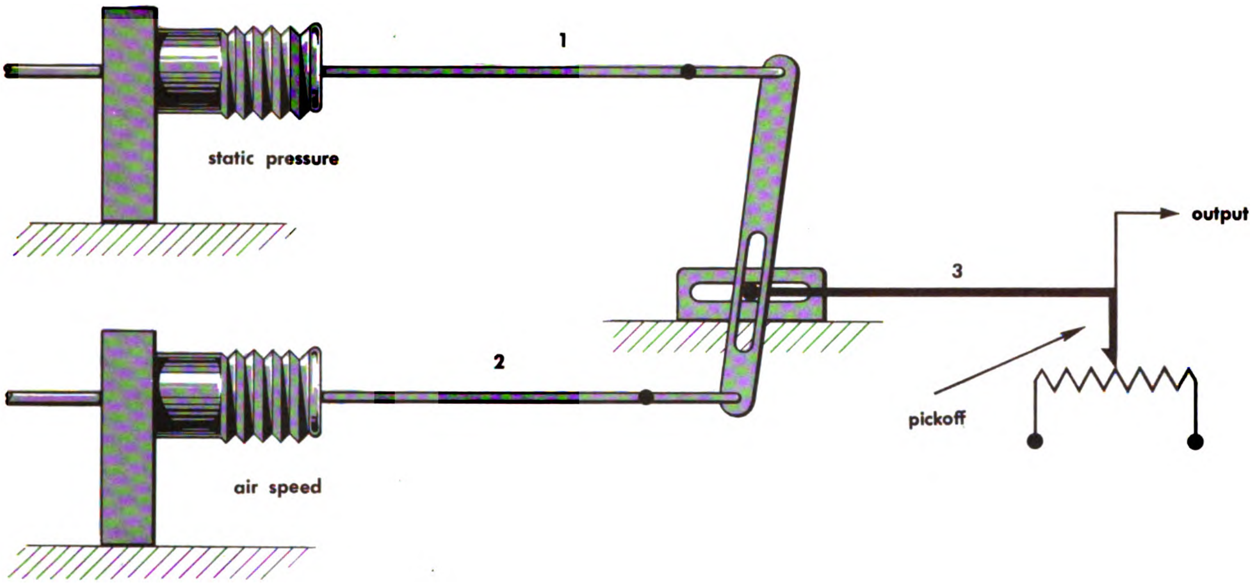
Differential Selsyns

So far, only the mixing of information represented by the amplitude of electrical AC or DC signals has been considered. Information represented by the angular position of shafts can be transferred and mixed by the use of differential selsyns.

A sample part of a control system using differential selsyns for mixing position information is shown in the figure below. The gyro locates the rotor of a standard selsyn. Since an excitation voltage is applied to the rotor, a magnetic field is created in a direction depending upon the rotor position. This magnetic field is converted to stator voltages so the field can be transferred to a differential selsyn. The direction of the magnetic field in the final selsyn depends on the original selsyn position *and* the position of both differential selsyn windings in relation to the final selsyn. The position of four devices has contributed to the input to the amplifier. The amount of influence from each device is determined by the mechanical linkages. The sense of the position information can be established by the mechanical and elec-



Combining information through the use of standard selsyns and differential selsyns



Mixing of signals by mechanical means

trical connections to the selsyns and differential selsyns. Any reasonable amount of differential selsyn windings could be included in a circuit since the weight and power losses are the only limitations. In the illustration, the selsyns are positioned to provide a null (zero output) to the amplifier.

Mechanical Mixers

Information can be combined by means of mechanical mixers which are made up of shafts and gears. Lateral positions can be combined by using plain levers as illustrated above. Assume that shafts number one and two operate independently and that the positions represent information which must be combined. The three connections pivot freely. The position of shaft number three represents a weighted average of the other two shafts, because vertical lever arms from shaft number three are not the same length. The sense is obtained by the direction of shaft movement.

The output from shaft number three could operate an electrical pickoff. The lever mixer would then eliminate the need for two pickoffs for converting the shaft positions to an electrical signal before mixing. If the inputs to shaft one and two provide enough power, the output could operate a pneumatic or hydraulic valve directly.

Gears can be used to combine position or angular velocity information by means of a standard gear differential of the same type as used in the rear axle of an automobile. If the input shafts contain position information, they will move slowly and maintain approximately the same average position. The position of the output shaft constantly indicates the difference between the two shaft positions. If the information is represented by the speed of rotation of the shafts, the angular velocity of the output shaft represents the difference between the input velocities.

The input shafts could be selected so that the output represents the sum of the inputs rather than the difference. This summation of inputs applies to both position and angular velocity information. The weighting factors could be controlled by changing the gear ratios.

Pneumatic and Hydraulic Mixers

Information sometimes is transferred by means of air or hydraulic tubes. Signals are created by the varying pressure within the tube. Two such signals can be combined by a union of the two tubes into one. An air pickoff controls the amount of air flow. The ratio of signals depends on the adjustment of the pickoff or the valve adjustment of the air pressure prior to the pickoff.

With this brief mention of pneumatic and hydraulic mixers, we end our considerations of mixers and move on to integrators.

HOW DIFFERENT INTEGRATORS FUNCTION

An integrator is part of the computer section of the control system since it performs a mathematical operation on an input signal. The integrator changes the input to a form that represents new information which is desired for one of several reasons presented later.

Simply stated, the integral of constant signal is proportional to the amplitude *multiplied* by the *time* this signal exists. For example, suppose the output of an integrator is four volts, and it was produced by a constant input signal which lasted for one minute. If the same input signal had lasted for only *one-half minute*, the output would have been *two* volts. Of course, this is assuming the integrator is a perfect one. The control integrator, then, is a device for computing the *time* that an input signal exists. In contrast, you will recall that the output of a sensor which provides continuous control is normally instantaneously proportional to some missile error. This proportional signal provides one fundamental means of continuous control. Integral control is another method of providing continuous control.

Normally, an actual missile error is not constant as assumed in the above example. Instead, the amplitude and sense of the error changes depending on missile flight conditions. Even so, the correct integration is proportional to the product of the operating time and the *average* error during that time. If the sense of the error should change during the integration period, the signal of opposite sense would decrease the final output of the integrator. The integrator can be considered to be a continuous computer since it is always producing a voltage which is proportional to the product of the average input and the time.

It can be stated that the integration of an error with respect to time represents a summation or accumulation of error over a specified period of time. Actually, one definition for the term *integrate* is: "To indicate the whole of, to give the sum or total of."

Any integrator possesses a lag effect. Assume that a constant error is applied. Although the input signal suddenly reaches a certain value, no time has elapsed, so the output is zero at that instant (assuming the output originally nulled). The output will steadily increase as in the previous example. Therefore, the reaction is considered to be somewhat delayed.

The graphs for constant error inputs and the output, assuming the integrator itself operates without error, are shown on the following page. The lag effect can be observed since time is involved before the output presents an appreciable signal. Time is also required for an appreciable reduction in the output after the input is removed.

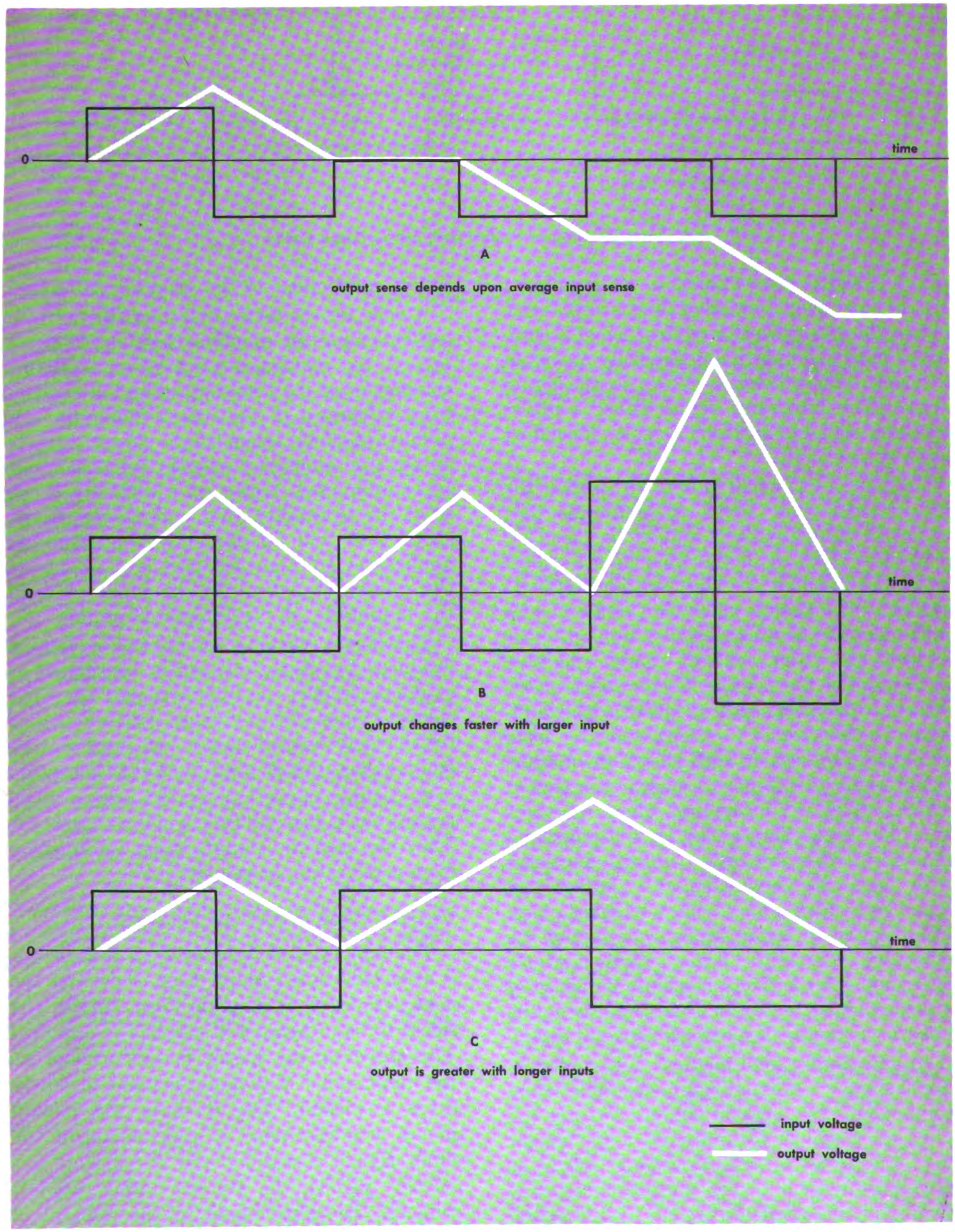
The outputs from actual integrators may deviate from the outputs shown in the figure. Such deviation would be due to characteristic errors in the particular integrating devices.

An integrator also has a filtering effect on relatively high frequency signals. Suppose a more rapidly varying signal with the same amplitude is applied to the input. The output would indicate practically nothing since the period of the input signal is so short. The device filters high frequency signals since it is relatively insensitive to rapid variations in the error signal.

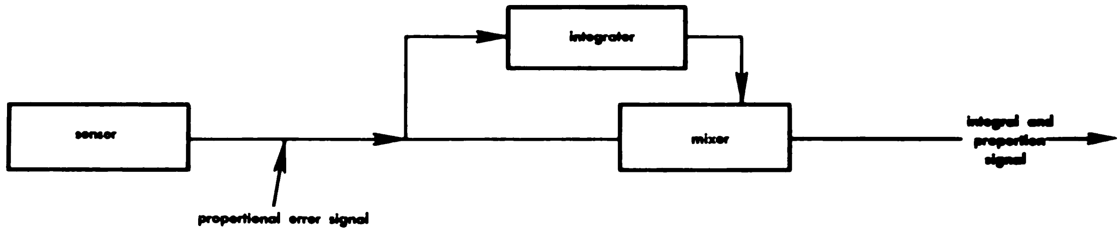
Some types of integrators are more accurate than others by virtue of the principle of operation. Several problems are involved in producing an integrator which will deliver an output exactly proportional to the integral of the input. First, the integrator must be able to react linearly to any signal level which may be applied to it. That is, if the input signal becomes doubled, the output of the integrator should increase at twice the previous rate. Second, this reaction should occur rapidly. Third, the integrator must have the ability to "remember." Suppose a certain signal is developed at the integrator output and there is no further input signal. For perfect results this same signal should remain at the output. The above three qualities are present to different degrees in integrators, depending mostly upon the type used.

Uses of Missile Integrators

There are situations in missile systems in



Perfect integrator outputs



Inclusion of integral control with proportional system

which a signal proportional to an accumulation of error over a period of time is important. Normally this integrator control signal is combined with a proportional signal by means of a mixer. This process is shown in the figure above. The integrator signal either *supports* the error signal or *subtracts* from it, depending upon the purpose.

An integrator signal is used to *support* the proportional error signal to guarantee that sufficient correction will always be made by a system. Such an application appears later in the discussion of the hydraulic-electric system. The degree of control that a pure proportional signal can exert is limited. Over control (or under control) causes excessive movement of the missile. There are instances when proportional control alone is not sufficient to overcome a strong, steady force deviating a missile. In such a case, the proportional error signal has a steady component which affects the integrator. Since the error is of one sense (determined by the crosswind direction), the integrator output increases with time and augments the proportional signal until correction ultimately takes place. The integrator output remains to supply the necessary correction signal component to overcome the relatively constant influence of a steady cross wind. In this case, then, the integrator supplies a necessary constant signal.

An integrator signal *subtracts* from a proportional signal to eliminate an undesirable signal component. The general connections are the same as shown in the previous figure except that the sense of the integrator output is reversed. As stated earlier, the output responds to long-lasting errors but is relatively insensitive to rapid variations. When the proportional and integral signals counteract, the result is to reduce or eliminate an undesirable steady component of the proportional signal.

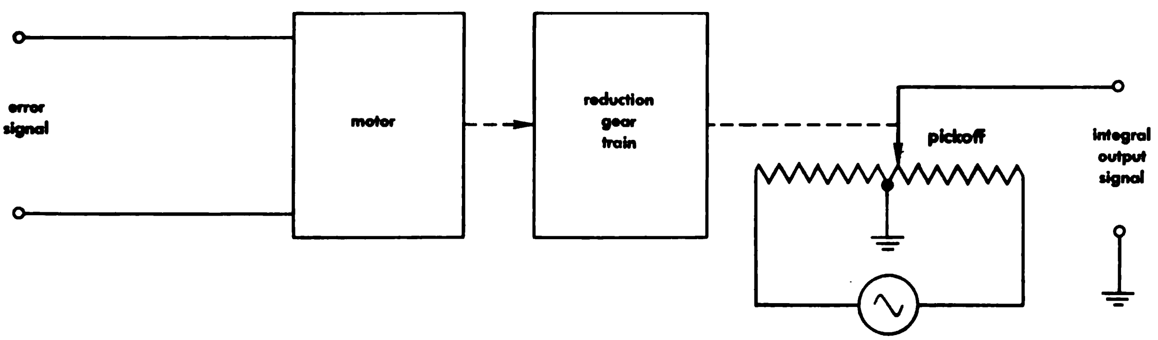
A need for the latter characteristic exists in self-balancing an amplifier output. A signal caused by unbalance appears as a steady component of the total output and is opposed by the integral of the input. Similarly, any unbalance which arises from a sensor, such as autosyn misalignment or gyro unbalance, can be counteracted. However, a difficulty arises from the possibility that a desirable steady signal also will be eliminated. In such an application, sufficient feedback is allowed from the integrator output to the input to limit the amplitude of the integrator output.

The slowly varying component of voltage produced by the output can be considered as a reference for the more rapidly varying signals. Since this reference voltage is not restricted to a particular value, integral control is sometimes referred to as *floating control.*

A further important application of integrators is in inertial type guidance systems. They are used to compute range information from acceleration or velocity data.

Variable-Speed Motor Type Integrator

The first type of integrator we will discuss is a variable-speed motor type, shown on the next page in block diagram form. Briefly, the signal to be integrated is supplied to a motor. The speed of the motor is proportional to the input signal. The motor is made to slowly drive a pickoff. The distance the pickoff moves is proportional to the integral of the input signal. Assuming that the pickoff was initially zeroed, the voltage output from it will also be proportional to the integral of the input signal. The diagram shows such an operation using a linear potentiometer (pot) to detect the integration.



Variable-speed motor integrator

The error signal is first fed to a power amplifier so the signal will be strong enough to drive the motor at varying speeds. A large gear reduction follows the motor to prevent it from driving the contact to either end of the pot during normal integrator operation. The signal slowly increases with time as the potentiometer is driven further from the midpoint or balanced position.

The motor rotates in a direction dependent on the polarity or phase of the input signal. If the motor is AC, the integration is performed on the envelope of the input signal. In actual use, the error amplitude is not constant but varies irregularly. It may even reverse sense. Physically, this reversal would produce opposite rotation of the motor.

For simplicity, suppose the integrator suddenly receives a signal of amplitude "B." The speed of rotation of the motor is proportional to the input signal,

$$\text{or } V_M = K_1B.$$

Therefore, the contact slowly progresses to the right or left of the zero position at a speed proportional to the speed of the motor,

$$\text{or } V_C = K_1K_2B.$$

The *distance* the contact moves is the speed times the time:

$$d = V_Ct, \text{ or}$$
$$d = K_1K_2Bt,$$

where "t" is the time the constant error exists. The output voltage is proportional to the distance the potentiometer contact moves from the zero or balanced position,

$$\text{or } E_o = K_1K_2K_3Bt.$$

The three constants each depend on certain independent factors of the device itself such as motor rpm, gear ratio, and voltage applied to the bridge circuit. They can be lumped into one:

$$E_o = KBt,$$
$$\text{where } K = K_1K_2K_3.$$

This shows that the output voltage is proportional to the input voltage multiplied by the time. The constant of proportionality "K," depends on the machine and must be considered in the designing of the system.

In order for the above integrator to operate perfectly, two conditions must be met. The speed of motor rotation must be directly proportional to the input error voltage, and the output from the potentiometer must always be proportional to the distance the contact has moved from the zero position.

These two ideal conditions are difficult to meet in actual practice. For instance, the curve of input power versus motor speed of rotation is never exactly a straight line. An inverse feedback circuit helps to remedy this problem. The second problem requires the use of linear potentiometers. If a selsyn is used as the pickoff, error will be instroduced for large angles of deflection. The extremes of the particular pickoff establish the operating range of the integrator.

An integrator need not be perfect, however, for it to operate satisfactorily in a missile control system since it needs to operate only within certain limits. These limits are established during the designing of the control system; they depend on missile aerodynamics and on the control system itself.

Initial conditions. Normally, the pickoff which provides the output is zeroed before

the beginning of integration. Sometimes a beginning position *other* than null is required. The integrator would then begin with a constant output and any changes would depend on the input. Such a beginning bias signal is important since it represents a condition which should exist at the beginning of integration. It is called the *initial condition*.

Ball-and-Disc Mechanical Integrator

A ball-and-disc type integrator is used in precision calculating devices such as bombsights and analogue computers. This type is applicable to missile use. Its operation is comparable to the variable speed integrator in which the speed depends on the error signal. In this case, the motor speed is constant and the *gear ratio* is varied depending on the error signal. The ball-and-disc type integrator is shown in the figure below.

The ball is turned by contacting a rotating disc. The contact point of the disc varies from the disc center to the circumference, depending on the error signal. When the ball is at the center, it does not turn. It turns most rapidly at the outer edge. The distance the ball rotates is proportional to the integral of the error signal with respect to time. This distance is detected by some type of sensor.

The amplitude of input determines the distance "B" in the drawing. The speed of rotation of the selsyn pickoff is slow. It depends on the disc rotation speed, the distance "B," and the gear ratio. The selsyn deflection depends on the speed and the time. For small deflections from null, the selsyn output will be

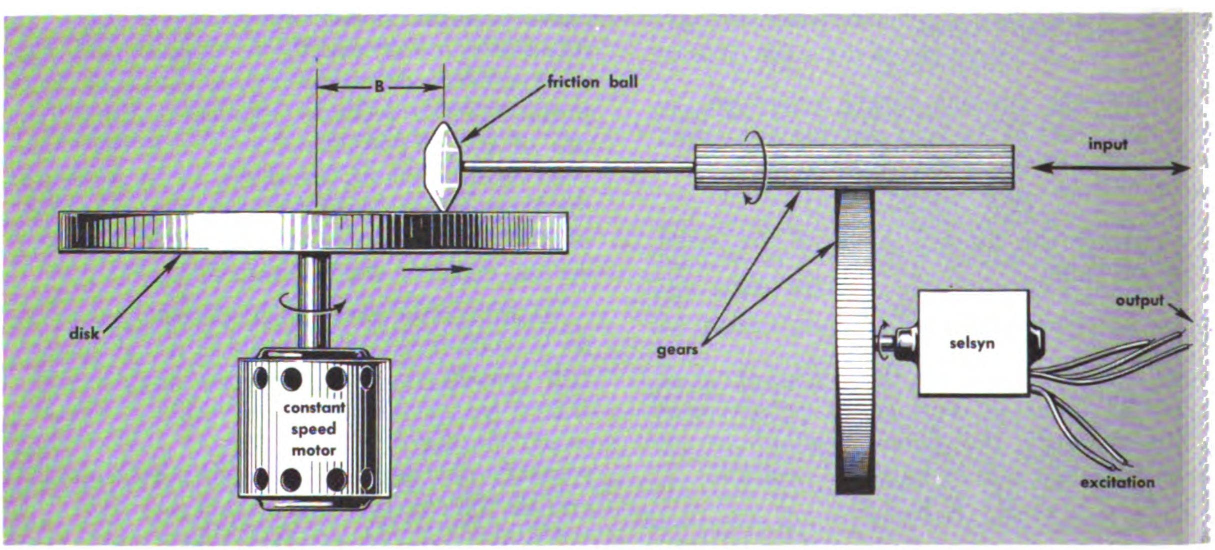
$$E_o = KBt.$$

This is the same result that is obtained from the variable speed integrator, although the constant of proportionality "K" may be different. As the expression shows, this voltage is again proportional to the integral of the signal with respect to time.

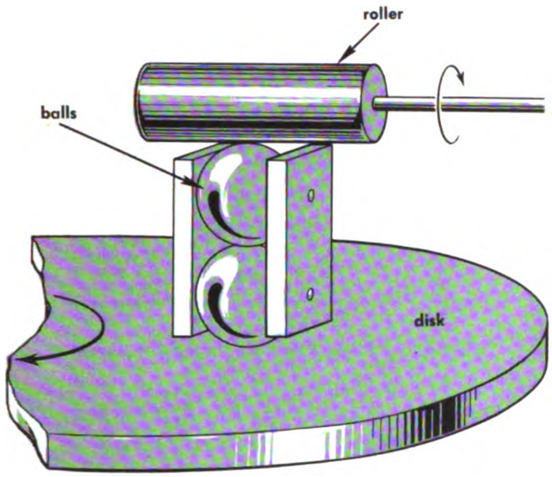
The expressions above assume no slippage between the ball and disc. Since slippage introduces errors, pressure exists at the contact point. Lack of precision and wear also introduces errors. The additional ball and roller shown in the drawing of a disc-contact system is used to overcome certain errors inherent in the fundamental integrator.

Any function in a missile is normally integrated with respect to time. A constant speed motor can set up a time reference in which a certain number of disc revolutions represents a certain time. Generally speaking, however, integration could occur with respect to any other reference. In such a case the speed of the disc is altered to represent the other reference.

Both the variable-speed and constant-speed motor integrator depend on the fact that a



Ball-and-disc integrator



Integrator disc contact system

displacement produced by some velocity function is proportional to the integral of the velocity function. This is supported by:

distance = velocity (average) × time,

where distance depends not only on velocity but on the time this velocity exists.

R-C Integrator

A simple and commonly used integrator consists of two circuit elements: a resistor and capacitor. The fundamental circuit of an R-C integrator is shown in the figure at lower right.

The voltage across a capacitor is proportional to the integral of the charging current. It can be explained by considering that the voltage across a capacitor is

$$E = \frac{Q}{C}$$

For any given capacitor (C), the voltage depends directly on the charge (Q) which is the unbalance of electrons on the two plates. The amount of this charge depends on the current flow *and* the time which this flow exists. If you will remember, these conditions are comparable to the definition of integration given on page 208.

The fact that the voltage is proportional to the integral of the charging current allows the R-C circuit to be used as an integrator. The capacitor voltage is the integrator output. Provision must be made to supply a charging current that is proportional to the input information. The purpose of the resistor is to produce this proportional current from an input signal voltage "e_i." At the instant this voltage is applied, the charging current becomes

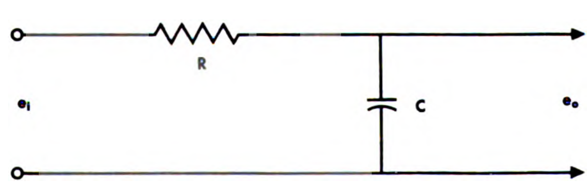
$$i = \frac{e_i}{R}$$

Unfortunately this proportionality does not continue to exist. As the capacitor becomes charged, the capacitor voltage opposes the charging current, and the charging current becomes less proportional to the input signal. The resulting error in the output is shown by the curve of actual "e_o" in the first figure on the next page. The ideal output, for a constant input signal, would be steadily increasing as shown on the graph. This steady increase is attained only when the signal voltage is first applied and the capacitor has not become appreciably charged.

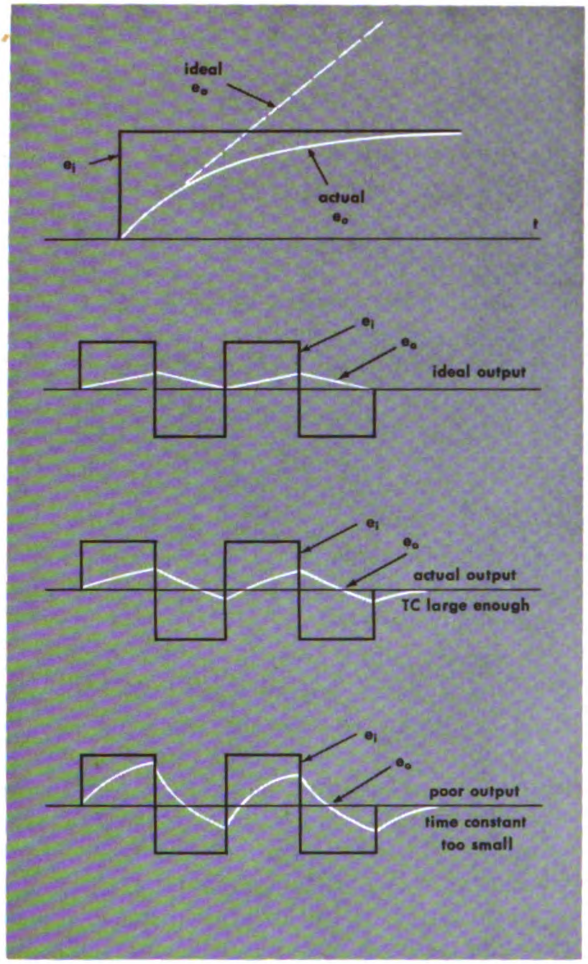
A remedy to this error in the R-C integrator is to use a circuit with a large time constant. Such a circuit delays the charging of the capacitor. The result is a more accurate integration of an input signal as shown in the two curves using a square wave input as an example. The ideal output would be a perfect triangular wave. Although a long time constant produces more accurate results, it also provides a much lower output for the same input signal. Better integration is possible by the use of a high gain, feedback amplifier.

Amplifier R-C Integrator

The common integrator shown diagrammatically in the lower left figure on the next page uses a high gain amplifier, with a capacitor forming a feedback path for output variations. This is the Miller integrator. The amplifier produces an output which is not



Fundamental circuit of R-C integrator



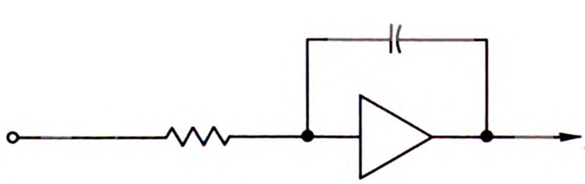
Curves showing error in simple R-C integrator

limited by the input signal as it is in the simple R-C integrator. The amplifier also supplies any energy which is required in the output. The function of the input signal is to control the charging current.

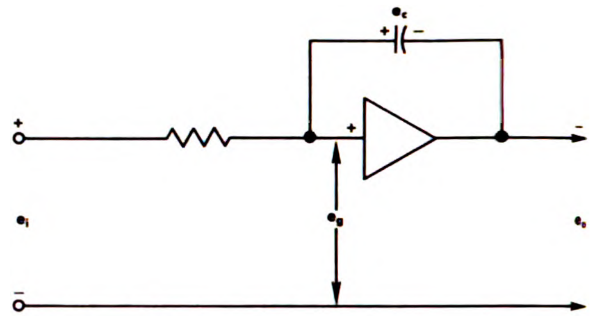
The operation can be explained by assuming a constant input as shown to the right above. At the start, assume the initial condition is zero; that is,

$$e_i = e_g = e_o = 0.$$

Also assume that the capacitor is discharged.



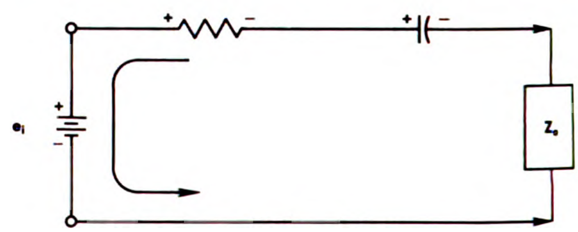
Amplifier integrator



Amplifier integrator with positive input

The positive voltage to be integrated, "e_i," is then applied. The capacitor charges with a polarity as shown, since electrons are attracted from the left plate. The charging path is shown below.

A voltage measured at the amplifier input, "e_g," tends to rise in the positive direction since this point is directly coupled to "e_i." However, this rise tends to be opposed by



Electron path for capacitor charging

the degenerative feedback voltage from the output. The output will be $-Ae_g$. The letter "A" stands for the amplifier gain. The minus sign indicates that the output polarity or phase is opposite to the input. The output changes "A" times faster or steeper than "e_g." The output voltage will be negative and will aid the charging of the condensers.

For a certain input voltage, the charging current is limited to a *particular* value which tends to keep "e_g" practically zero. If *the* current should *exceed* this value, "e_g" would decrease a small amount due to the increased voltage drop across R. Then $e_o (e_o = -Ae_g)$ would decrease, and the charging current would decrease to the original value. If the initial charging current should *decrease*, the opposite action would occur. The value of the charging current is therefore stabilized to a specific value proportional to the input volt-

age. This eliminates the error caused by "e_i" and the charging current not remaining proportional in the fundamental R-C integrator.

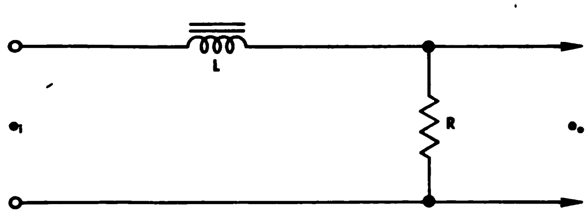
This constant charging current must be produced by "e_o" despite the fact that the steadily increasing capacitor voltage opposes the charging current. To do this, "e_o" must also steadily increase. This steady increase in "e_o" is exactly the integrator output voltage desired for a constant signal input.

Similar action would be produced for a condition in which the input signal suddenly became negative. Polarities then would be in reverse to those shown in the example given. Remember that simple examples are used for explanation on the assumption that the desired result also will be produced for a more complicated signal input. Removal of "e_i" would produce little effect upon the output which existed at that instant, since the amplifier output would oppose the tendency for "C" to discharge.

The limits for "e_o" are determined by the amplifier and not by "e_i" or the range of "e_R." The output range would be designed to produce an increasing output for any probable input amplitude and period of application. The exception to this would be an integrator which was designed to function also as a limiter.

R-L Type Integrator

A resistor and inductor combination can also be used as an integrator. Notice in the diagram that the components of the R-L integrator are connected in reverse to the R-C integrator. This circuit produces the same response as the R-C circuit. It also possesses the same defects. A further error exists because of the resistance in the coil. The R-C circuit is used more often than the R-L circuit.



Simple R-L integrator

Thermal Integrator

An integrator can operate on the principle that the temperature of a heating element is approximately proportional to the integral of the current input for small intervals of time. A heat detector could then provide the integrator output. Such an integrator would be limited in "memory."

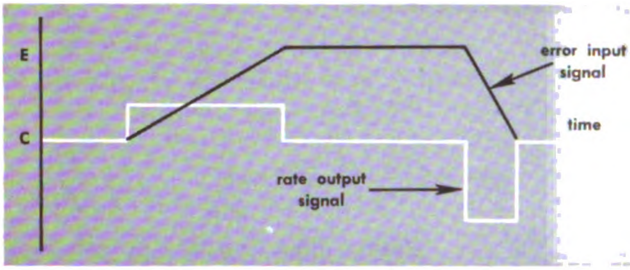
Integration From Viscous Damped Integrating Gyro

A method of integration is made possible by using a viscous damped integrating gyro, such as was discussed in the section on gyros. This gyro is not considered a free gyro since it has only two degrees of freedom. The integration function is produced as the gyro is allowed to precess through the restrictive force imposed by the oil. The force causing precession of the gyro is produced by angular deviation of the missile. This force depends on the rate of change of missile deviation. Precession of the integrating gyro is not limited to a few degrees as in a rate gyro. The amount of precession depends upon the force causing precession and the duration of precession. The missile deviation also depends upon the rate and time of angular motion. Therefore, a definite ratio exists between the angle of gyro precession and the angle of missile displacement. Although the force on the gyro is due to rate of displacement, this gyro produces displacement information as a result of the integration which takes place within the gyro unit.

We now are ready to take up rate systems, the last of the three general types of components of a computer unit.

FUNCTIONS OF RATE SYSTEMS

A rate system is part of the computer section of the control system since it, like the integrator, performs a mathematical operation on an input signal. This operation, however, is much different from that of the integrator. The input signal is changed to a new form which represents new information. This information improves the control system stability. The input signal, remember, conveys some type of missile information such as position error, angular deviation, altitude or air speed error, or control surface position.



Rate output depends on rate of change of input signal

Briefly, a rate component is a device which produces an output that is proportional to the *rate of change* of the input signal amplitude. This mathematical operation that determines the speed at which a signal is changing is called differentiation. If an input signal is increasing or decreasing slowly, the output of the rate component is small. If a signal is changing rapidly, its output is proportionally larger. Even when an input signal is large, the output of a rate component should be zero so long as the input signal remains steady. The output of most rate components reverses sense when the input changes from an increasing to a decreasing signal. The illustration above shows rate output from sample inputs.

You probably have more contact with rate components (differentiators) than you realize. Any device which indicates the speed of some object directly is a rate computer. The automobile speedometer is an example of this. The input of the speedometer is a rotating shaft. While the mileage section is simply an extremely large gear reduction, some sort of rate sensing device is necessary to detect and indicate speed.

This rotational velocity can be detected by centrifugal force or by magnetism. Suppose a weight is fastened to a rotating member. The weight is fastened so that it can swing outward against spring tension, The amount of centrifugal force tending to pull the weight from the shaft depends on the speed of rotation. Hence, the displacement of the weight is a measure of speed.

A special generator could be used to indicate rotational speed since the rate at which magnetic lines of force are cut determines the output voltage. A voltmeter would then be calibrated in rpm, miles per hour, or comparable units.

Purposes of Rate Control

A correct answer to a question regarding the purpose for rate components is "To improve stability." This answer is general, however, since it could apply more or less to every device in a control system. Rate systems would not be necessary were it not for the fact that a definite amount of time is required to perform any type of control operation. In the case of a missile, the correction of missile deviation cannot occur at the instant of deviation, for this would result in no deviation whatsoever. Such a result would be ideal, but it is impossible since a system must detect some error in order to operate. The degree to which this ideal situation is approached depends on the design of the system and the requirements of the load such as size of airframe and missile aerodynamics.

Since the ideal control system described would have zero time lag and would result in a missile which did not deviate whatsoever, efforts are made to reduce this time lag to a minimum. The airframe and control surfaces are designed to correct missile attitude rapidly. The control surfaces are moved by powerful, fast-acting actuators using an amplified error signal to produce fast response. However, the reduction of lag by such measures approaches a limit.

OSCILLATION FROM OVERCONTROL. At this point, you may be thinking, "Why not increase the amplification of the error signal from the sensor so that even a small missile deviation will produce a signal to correct the missile almost instantaneously." This idea is all right to a certain extent. A signal which is *too* large, however, returns the missile beyond the desired point to an opposite error. The resulting opposite error signal causes the missile to deviate back to the first direction. The end result is a serious oscillation about the desired condition.

DAMPING UNWANTED OSCILLATION. The addition of a rate signal has the effect of *damping* the oscillation to a certain extent. As you know, the term "damp" means to restrain or retard. The amount of damping is classified as critical-, under-, or over-damping.

Rate Component Creates Lead

If the order signal which commands system action could be advanced in time, the effect would be an apparent reduction in the overall lag of the system; a rate signal does this. It is combined with the proportional signal to produce a resultant signal which leads the original proportional signal.

The combination of the error signal with a rate signal to produce a leading signal is shown on the following page. The proportional signal is assumed to be a sine wave as shown at the top. If it represents yaw error, this indicates that the missile is "fishtailing" to either side of the desired heading. A sine wave is used since it is most suitable for an example, despite the fact that it actually represents an undesirable flight condition.

The second waveform shows the output of a rate device when using a part of the proportion signal for an input. Notice that the output of the rate device is zero at the instant the input signal is at a crest and that the output has no rate of change. The error curve has the steepest slope when the input error signal is changing at a maximum rate. This phenomenon occurs when the error voltage is zero. At this point, the rate signal is a maximum, indicating a maximum rate of change.

The vector addition of two sine waves (rate and proportion) which are out of phase produces a sine wave with a phase somewhere in between. This result is shown in the third line of the illustration on the preceding page.

The two signals are mixed in a ratio which produces the best response from the system. This optimum ratio is determined by the other system components and the aerodynamics of the missile, it may vary with any change in missile configuration, weight, or altitude. Of course, the amount of rate signal needs only to be within a certain range for acceptable missile operation.

Remember that there is a rate output only when the missile deviation is *changing*. If the rate signal were not combined with the error signal there would be no way for the control system to respond to a constant error. It is also possible to combine an attitude rate signal with a guidance error signal.

Predictor Circuit

A rate circuit is often called a lead circuit, phase-advancer, differentiator, or prediction circuit. The terms are analogous and vary only in the concept of the subject. The term "rate" refers to the fact that the signal is proportional to the rate of change of the input signal. The term "differentiator" refers to the mathematical operation which is performed on the equation of the input voltage to yield the proper equation for the output voltage. The illustration on the next page shows the origin of the terms "lead circuit" or "phase advance."

But, why call a rate device a "predictor?" Consider the heavy lined portion of the proportional signal in the illustration. The error is increasing during this period. The resultant during this time is much greater than the value of the proportional signal alone. If the missile should *slowly* veer from the desired heading, the resultant signal would be due mainly to the proportional signal.

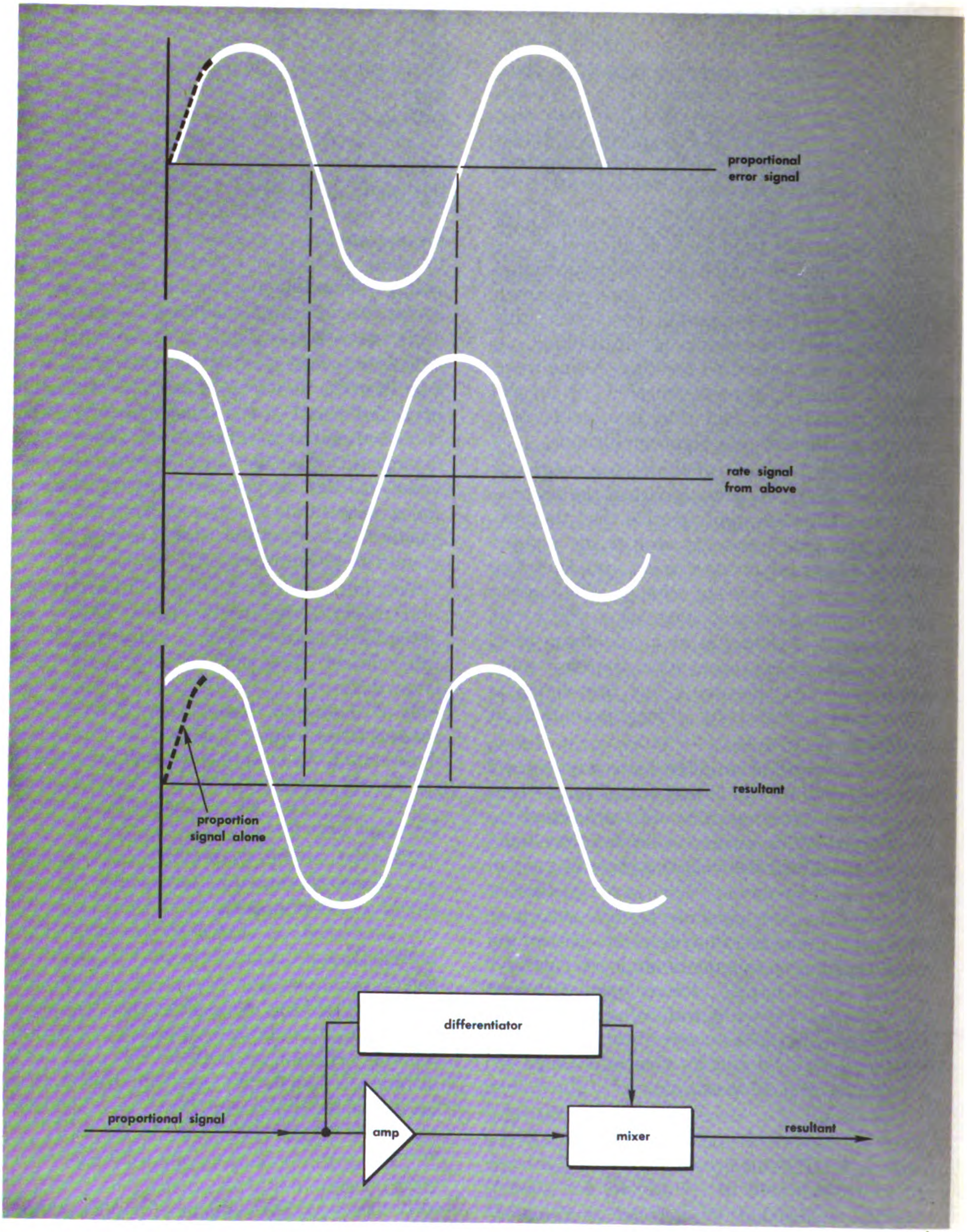
Without a rate circuit, a rapid missile deviation would normally produce a fairly large deviation before correction takes place. The rate circuit "predicts" this possible future large deviation from the existence of a high *rate* of deviation, and it increases the corrective effect. The result is that the missile does not deviate as far with a rate unit in its control system.

Electronic Rate Circuits

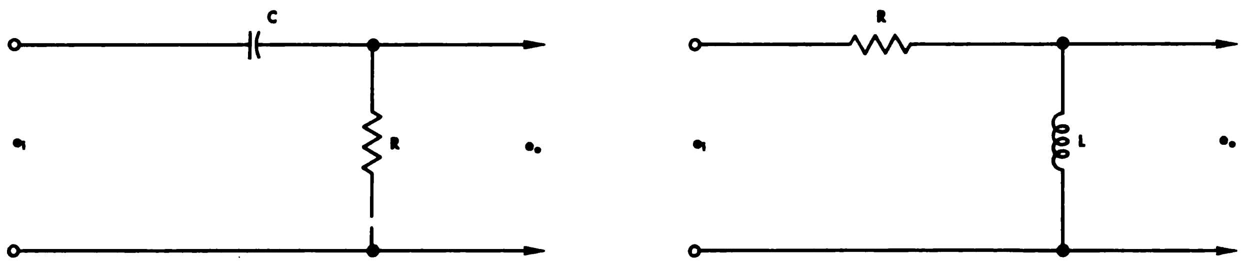
The R-L and R-C combinations which are used as electronic integrators also serve as effective differentiators when the components are connected as shown in the illustration, Differentiator Circuits, on the next page.

The inductor circuit produces a differentiation because the voltage across the coil is proportional to the rate of change of current. This is assuming a perfect coil with no resistance. The IR drop across the coil causes an error in the output voltage. Therefore, the R-C circuit is used more often.

The capacitor circuit produces a differentiation because the charging current is proportional to the rate of change of applied voltage. The output voltage dropped across the resistor in the circuit is dependent upon the charging



Method of producing an error signal lead by differentiation

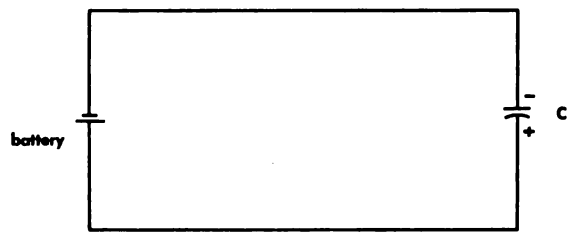


Differentiator circuits

current. Think back to the basic capacitor theory in which the voltage across a capacitor is determined by the unbalanced electrons on the two plates. In the circuit at the right, the battery transfers electrons from the lower plate to the upper one, so "e_c" equals the applied voltage. Now, suppose the voltage from the battery could be slowly increased at a steady rate. The charging current would be proportional to the *rate of change* of the voltage since "e_c" also changes. A capacitor thus forms the basis of a rate circuit.

The theoretical circuit illustrated would operate as a perfect differentiator since it contains no resistance or other impedance. Such a differentiator is an impossibility, however, because the voltage source and wiring would possess impedance. Furthermore, some method is necessary to detect the charging current. Since the current is normally small, a resistance is placed in series with the capacitor, and the voltage across it is used as the output. This impedance creates an error due to the dissipation of energy in the resistance and the delay in the response of the output.

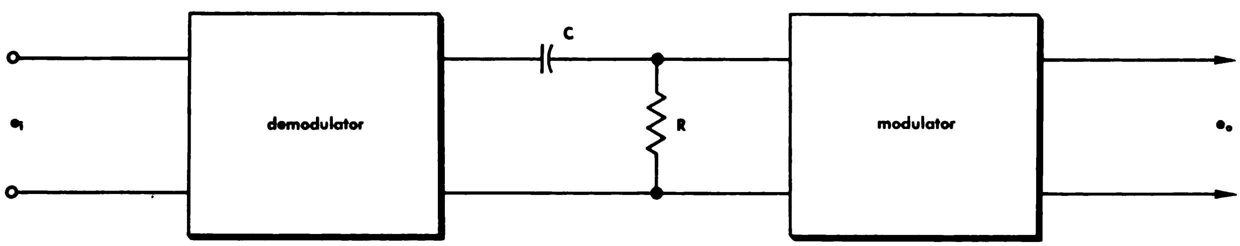
A delay exists because every RC circuit has a definite time constant. For example, if one rate of change input was suddenly changed to a different rate of change input, the output of the rate circuit would not immediately give a true indication of the new input. This error is caused by what are known as transient voltages. Use of a small R-C time constant reduces the lag in response. It also reduces the output.



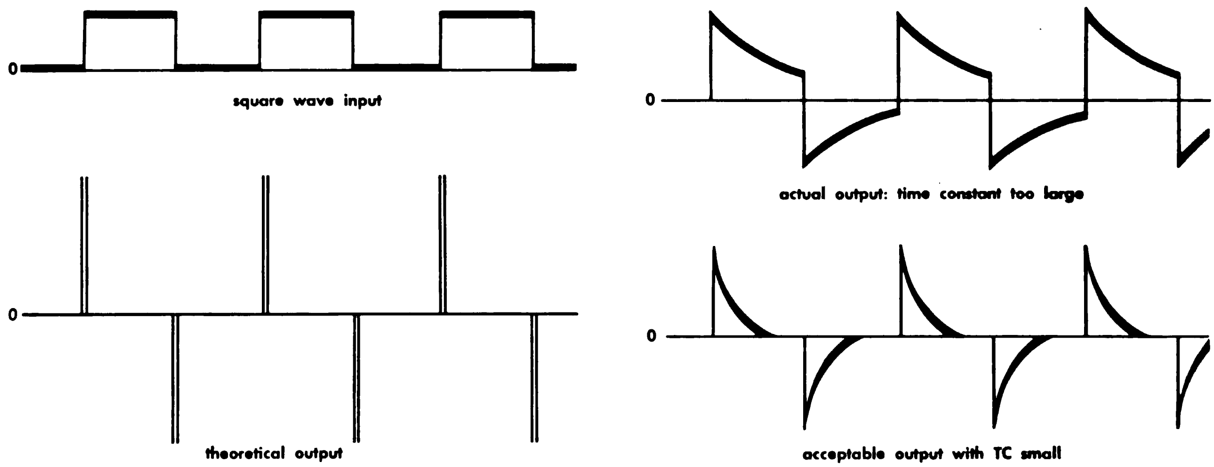
Theoretical differentiation circuit

In most cases a control system error signal is modulated by 400-cycle AC. This signal has to be demodulated to a varying DC voltage before it is applied to a capacitor rate circuit. This demodulation is necessary because the rate circuit is intended to detect rate of change of amplitude represented by the envelope of the input signal rather than the 400-cycle variations. The signal may be modulated again after it passes the rate circuit, as shown below. Modulation and demodulation also can be considered a computing function.

The first figure on the following page shows the rate output from a square wave input. A



Rate circuit for modulated signal



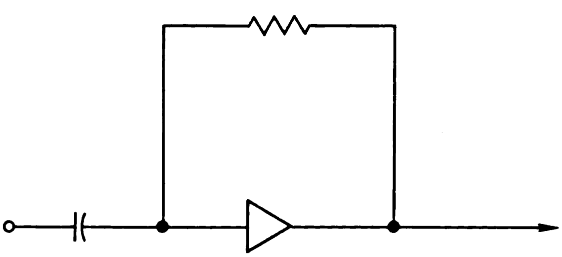
RC differentiator outputs

fairly accurate output is obtained if the time constant of the circuit does not exceed four times the period of the input signal. The perfect output is theoretical. It would be infinitely large when the input signal changes instantaneously. In actual practice this output is limited by the resistance in the circuit.

A solution to the errors in the basic R-C differentiator is to use an amplifier as was done in the case of the integrator. The amplifier supplies most of the dissipated energy and permits the use of a short time constant by amplification of the output.

The R-C Amplifier Differentiator

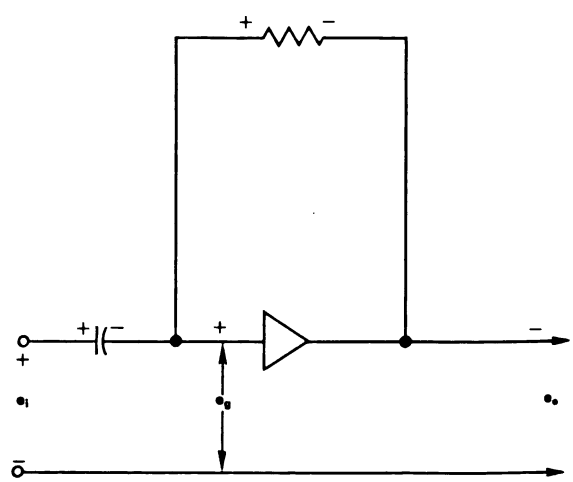
As shown in the diagram below, an R-C amplifier differentiator is just opposite to an electronic integrator. This differentiator circuit has a capacitor input and resistive feedback. The amplifier can be a modulator type so as to reduce drift and noise.



R-C amplifier differentiator

In analyzing such a circuit, it is normal to consider the amplifier to have infinite gain. On this basis, "e_g" can be considered to be zero. Although infinite gain is impossible, the amplifier does have a very large gain. If "e_i" should change, "e_g" would *tend* to change, but the inverse feedback voltage would practically counteract the change in "e_g."

The output voltage, then, would depend on the tendency of "e_g" to change. It would



Same R-C amplifier differentiator with positive going output

therefore be proportional to the rate of change of "e_i."

This same analysis can be given in terms of polarities. The preceding diagram shows the same circuit assuming a steady, positive input signal. "C" is charged to the steady value of the input signal, therefore,

$$e_g = 0$$
$$\text{and } e_o = 0.$$

Now, assume the error signal suddenly increases. "C" begins to lose electrons on the left and gain on the right. The charging of "C" tends to make "e_g" go positive. This action is reflected in the output as a negative voltage which increases at a rate "A" times faster than the input. Current produced by the output rapidly charges "C" to the new "e_i." The output is a large pulse of voltage which indicates the rapid change in the input voltage "e_i" This is the output desired. The amplitude of "e_o" is not limited by "e_i," as was the case in the simple RC circuit, because of the energy supplied by the amplifier. If "e_i" should decrease, an output pulse of opposite polarity would be produced.

Mechanical Differentiator

The ball-and-disc integrator, described earlier in the discussion on integrators, can be connected in such a manner as to produce an output proportional to the rate of change between two variables. A gear type mixer (gear differential) is necessary in the hookup. The actual operation of this mechanical differentiator is beyond the scope of this manual. The device is more pertinent to analog computers since the R-C amplifier differentiator provides sufficient accuracy for missile purposes.

Rate Gyro

A common method of producing a rate signal is by means of a separate sensor. A special type gyro, *rate gyro*, has been designed for this purpose. A rate gyro has one degree of freedom, but it is allowed to precess a few degrees in a second plane. Precession of the gyro in this plane is restrained by a spring which tends to return it to the midpoint. Any precession in this plane is a result of force on the gimbals. This force is caused by angular movement of the airframe about a certain axis and is proportional to the rate of deviation. The amount of small gyro deviation depends on the precessing force which must overcome the restraining force.

This gyro deviation can be detected with a pickoff, and it becomes a rate signal. A separate rate gyro is required for every plane of missile movement in which a rate signal is desired. These rate signals are normally combined with the output from a displacement gyro. Rate gyros were covered in the section on gyros.

RECAP OF COMPUTER FUNCTIONS

As you know, rate devices produce a method of continuous control in a system. They comprise the third method of continuous control that we have discussed. The three methods are:

Proportional
Integral
Rate

Proportional control operates the load by means of an error signal which is proportional to the amount of deviation of the controlled item from the desired point. This signal is produced by a sensor. Integral control operates the load by means of a signal which represents an accumulation of error. The integral signal is produced by devices covered in the discussion on integrators. Rate control operates the load by means of a signal proportional to the speed at which the deviation is changing. When rate or integral control is used, the output is usually combined with a proportional signal to produce the desired effects on the missile. Although rate and integral control have a cancelling effect, they sometimes appear in the same system. Practically every system contains a rate device since it enables the missile to react rapidly to any error which develops.

The computing operations need not occur in the position which is indicated by the basic block diagram. Regardless of where the operations take place, the mixing of signals and the development of integral and rate signals comprise the computing functions of a control system.

Reference Units of Control Systems

For our purposes, we will divide the mediums for control system references into three categories. These categories are *voltage references*, *time references*, and *physical references*.

Before taking up voltage references, the first category we'll discuss, let's take another look at the basic missile control block diagram shown below, so that we can see where the "reference" block fits into the system as a whole.

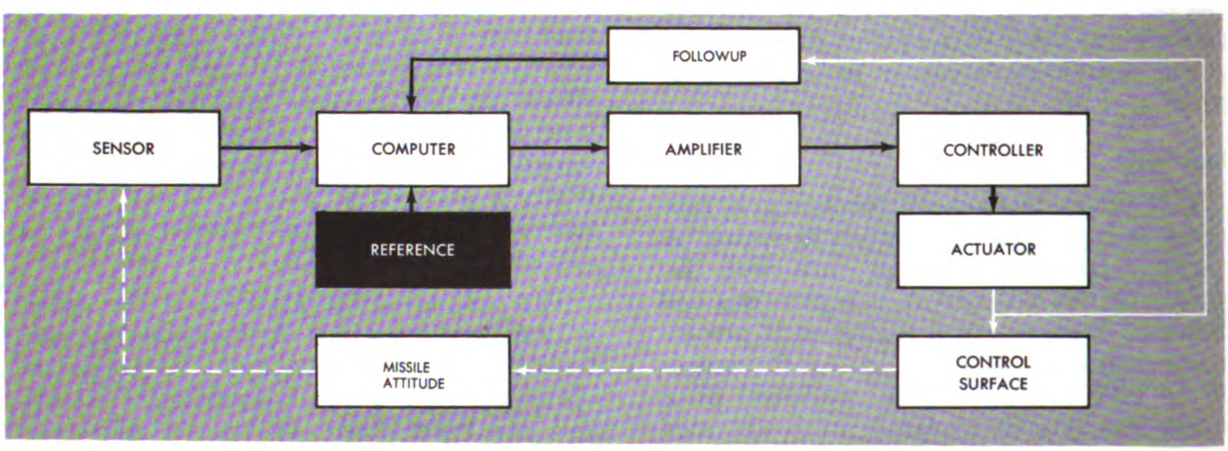
VOLTAGE REFERENCES

If there were no reference in a control system, the system would not know how to correct for an error. Consequently, an electronic control system could not function properly without some type of voltage reference.

AC Voltage Reference

In many of the existing control systems, the error signals are in the form of AC voltages. These error signals contain two characteristics which are necessary in order for control systems to make proper corrections. These characteristics are: (1) *amount* of deviation, or magnitude of the error, and (2) *direction* of deviation, or the sense of the error.

The *amount* of deviation is carried by the error signal in the form of amplitude. As the amount of deviation increases, the amplitude of the error signal also increases. If the amount of deviation should decrease, due to a movement of the control surfaces, then the error signal would decrease in amplitude.



Basic missile control block diagram

When the deviation decreases to zero, the amplitude of the error signal also decreases to zero. Therefore, there is no error signal when there is no error in missile attitude.

The *direction* of deviation is carried by the AC error signal as a function of phase. As has already been pointed out, this phasing, by itself, cannot tell the control system which way to correct unless there is something to which this phasing can be compared, such as an AC *reference voltage*.

Around *each* axis of control there can only be two directions of deviation. Thus, only two phases are required to tell the control system which way to move the control surfaces for any one axis. Usually one of these phases is in phase with the AC reference voltage, called *zero phase*, and the other is 180° out of phase with the AC reference voltage, generally called π (pi) *phase*.

By using phase sensitive circuits which compare the phase of the error signal with the AC reference voltage, the autopilot can remove the "direction of deviation" information from the error signal and move the control surfaces in the right direction to correct for the attitude error.

This AC reference voltage is usually the AC power supply for the control system. It also serves as the excitation voltage for the sensor unit which originates the error signal.

DC Voltage Reference

Most of the mechanisms used in the *controller unit* require a DC error signal, so it is necessary to show how a DC voltage can carry both of the characteristics that existed in the AC error signal. The *amplitude* of the DC error signal indicates *amount* of deviation. The *direction* of deviation is shown by the *polarity* (or in a broad sense, the phase) of the DC error signal.

The phase-sensitive circuits which are used to remove the two characteristics from an AC error signal usually produce a DC error signal that carries the two characteristics in the manner described in the above paragraph.

In some control systems, limiter circuits are used to prevent the error signal from becoming too large and causing *overcontrol*. These limiter circuits require DC voltage as their reference. This DC reference voltage can be obtained in many different ways, but regardless of the method used, the circuit functions as a part of the reference unit of the control system.

TIME REFERENCE

Time is another medium by which an autopilot can obtain a reference. Primarily, this reference takes the form of fixed periods of time during which certain autopilot functions are allowed (or not allowed) to take place. A definition of *time reference* as it applies to control systems is:

A time-delay device used to control a specific autopilot function during a predetermined period of time.

Characteristics of Timers

Timers vary greatly in their physical characteristics and their operation. Some of these physical characteristics need to be mentioned at this point to facilitate a better understanding of specific timers and their applications. These characteristics will be covered again as the various timing devices are discussed.

All control timers require some method of being initiated or triggered. Because all of the timing devices used in one system are not triggered at the same time, each must have its own trigger.

The method most used to trigger timing devices employs some form of electrical signal. In some cases this electrical signal is fed to a solenoid that has a core which moves when the solenoid is energized. The movement of this core, in these cases, is used to mechanically trigger the timing device. A special application of this method is in the case of the solenoid actually being a part of the timing device, rather than the triggering device. In some special cases the normal electrical signal to a time delay device is a voltage input, and the triggering takes place by removing the voltage.

There are two other electrical triggering methods. One is accomplished by applying the electrical signal to a heater coil which heats a thermosensitive bimetallic strip. The other electrical method is to apply the electrical signal to an electric motor. This signal

operates the motor, which is a part of the timing device. The voltage can be either AC or DC, depending on the type of motor used.

The only other method of triggering timing devices which has been used in missiles is to use an inertia switch. This method is discussed later under the piston-type pneumatic timer.

The need for a certain degree of accuracy greatly affects selection of a timer for a specific application. Some of the uses of timing devices permit use of timers having comparatively poor accuracy. Other uses, however, might require the highest degree of accuracy that can be obtained from timing devices.

When using timers in guided missiles, the effect that large changes in temperature and pressure have on their accuracy must be considered. Varying input voltage could affect the accuracy of some timers that make use of electric motors. These accuracy considerations on specific timers are discussed further when the timing devices are described.

The simplicity or the complexity of a timer is influenced by a characteristic which could be called the "repetition of operation." If the timer is to be used only once during a missile flight, usually the mechanism will be simple in operation. Conversely, if the timing device is to be used many times during a missile flight and has to recycle itself each time it performs its function, the mechanism usually will be complex in operation. The one outstanding exception to this is the thermal timing device which recycles itself without any additional mechanism.

In most timers the output produces electrical contact switching. The use of the timing device determines whether the contacts will be normally open or normally closed.

If the timer performs more than one function, there are two basic ways to accomplish the circuit changes. The first way is to employ a multicontact switch on the timing device; consequently, this method can close or open many circuits. The other way is to use a single-pole, single-throw (SPST) switch on the timer to energize or deenergize a relay that has the required number of contacts.

Applications of Timers

In the following paragraphs, some of the uses described might not be applied strictly as control system functions, but the applications are related enough to justify their mention at this point in the text. You might consider some of the applications as preset guidance functions and some as launch or power plant functions. The reason for mentioning such uses is to show the wide variety of applications of control timing devices.

One of the first uses of timers in guided missiles was to keep the gyros caged during the period of great acceleration at launch. The timer used for this purpose could be triggered by a voltage either from the ground equipment during the launch preparations or from an inertia switch which would close just after the missile leaves the launcher.

Timers can be used to trigger other timers, such as in the JB-2 (buzzbomb). In this missile, a mechanical timer was used to trigger another mechanical timer at a preset time after launch. The second timer was used to precess the directional gyro and make a preset turn. The length of time set into the second timer controlled the amount of turn.

In many cases timers are used to precess gyros and perform preset turns. Usually these turns are around the yaw axis, but in some of the terminal guidance systems, the gyro is precessed about the pitch axis to put the missile into a dive. The timers used to precess gyros have to be accurate.

Timing devices are used extensively in programming takeoff. Some timers will cut off the power plant if the boosters haven't fired within a given time after the power plant has developed some specified speed or thrust. This provides a safety period for clearing malfunctions. The sequence of firing the boosters is frequently controlled by timers. Launch programmers which control the climb of the missile often use timing devices.

Another possible use of timers is their application to command guidance. A carrier relay, which remains energized at all times during which the command carrier frequency is present, controls a timer. When the carrier frequency is lost, the carrier relay triggers the timer. After a certain time delay, if the

carrier has not been received, the timer completes its function and allows some preset command guidance function to take place. The Air Force Missile Testing Center used this application of a timing device in testing a missile. The device put the missile into a turn until the command control carrier frequency was again received by the missile.

One other specific use of timers is in connection with gyro erection and slaving systems. During a time delay, set up by a timer, the gyro is precessed at a much faster rate than under normal operating conditions. After this period of time has elapsed, the circuitry is set up by the timer to allow the gyro to precess at its normal rate.

Mechanical Timers

The only mechanical timer which has been used in a missile system is a clock mechanism. In operation it is similar to a mechanical alarm clock. The ordinary alarm clock can be set for a time delay up to twelve hours, while the clock mechanism used as a timer usually can only be set for relatively short periods in the order of seconds or minutes. The alarm clock gets its movement from energy which has been stored up in the main spring by turning a knob on the back of the clock. The clock mechanism timer gets its movement in much the same way. In order to set in a time delay, the knob on the face of the timer has to be turned so that the arrow points to the desired time on the dial. This puts energy into the main spring of the timer and corresponds to winding the alarm clock.

A bell rings at the time set on the alarm clock. When the time has elapsed on the clock timer, a switch is closed (or opened according to the application of the timer).

The time delay set into a clock mechanism, may not be used until the missile is in flight, so some sort of triggering linkage is necessary. This linkage usually consists of a catch which can be released by energizing a solenoid. This type of timing device can be used only once during a missile flight. Its accuracy is considered good.

Two specific applications of the clock type timer were in the JB-2 control system, which is described in Chapter 7, Section 1. In this system one five-minute clock mechanism triggers a sixty-second clock timer. The timer which can be set for any time up to five minutes is triggered at launch and determines when the preset turn will start. The timer which can be set for any time up to sixty seconds precesses the displacement gyro to the final course heading.

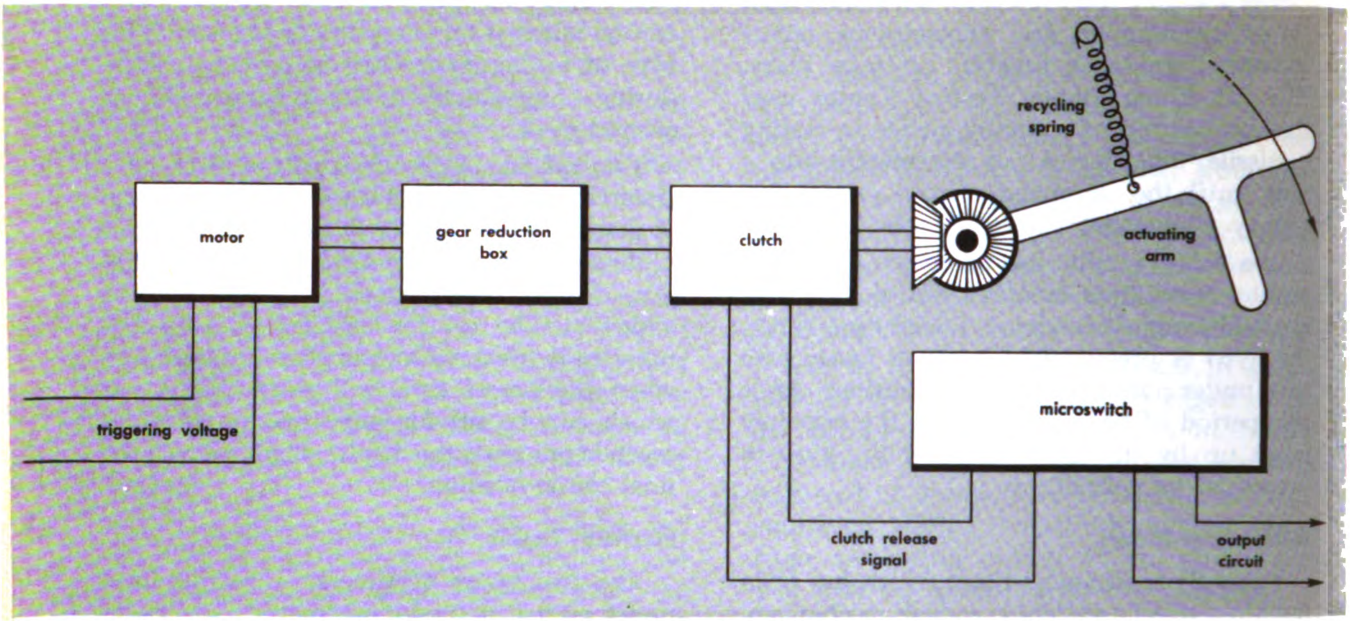
Electrical Timers

The two types of timers to be discussed under electrical timers are motor timers and thermal timers. In all cases the triggering is accomplished by an electrical signal. When the voltage is applied to the timers, the time interval or period of delay begins.

Motor timers. The simplest motor timer, shown in the accompanying illustration, possesses the operation principles of all motor timers.

A motor is turned by the triggering voltage. The speed of the motor's shaft is reduced by a gear box. The speed is such that the output shaft from the gear box does not make a complete revolution during the period of delay required. For example, if a thirty second delay were needed, the speed of the output of the gear box would be about one or two revolutions per minute. There is an arm connected to the output shaft of the gear box. The time it takes this arm to go from the starting position to a point which closes the switch is the time delay of the timing device. This simple mechanism can be used only once during a missile flight. If recycling is re-

Simple motor-timer operation



Operation of a recycling motor timer

quired during flight, the mechanism is more complex, as illustrated in the accompanying block diagram.

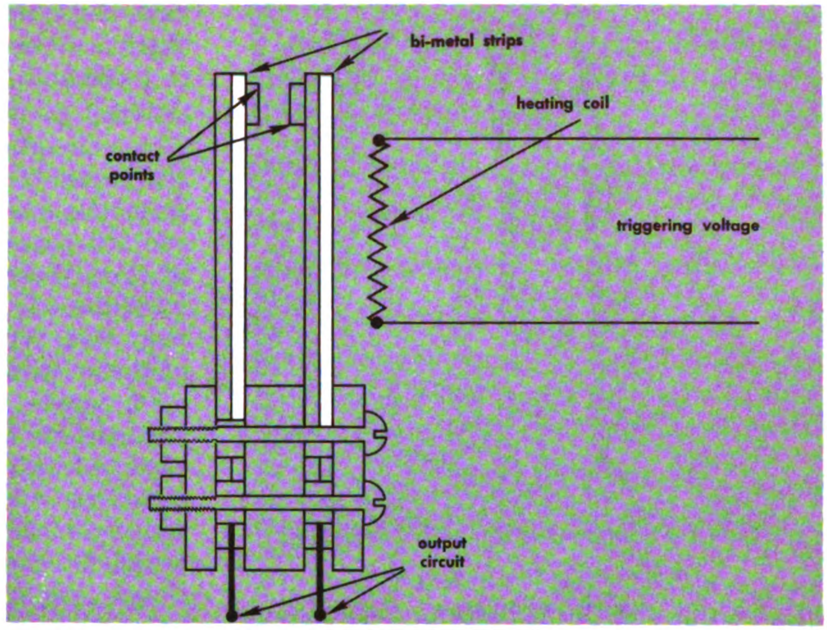
Again, the basic operational components are present: the triggering circuit, the motor, the gear box, and the switch for an output. But some additional stages have been used to make the timer automatically recycle itself. A clutch mechanism is necessary so that the linkage between the actuating arm and the gear box can be broken. The arm can then return to the starting position without the motor or gear box turning. The recycling spring pulls the actuating arm back to its original setting when the clutch disengages.

The microswitch, as shown in the diagram, has two sets of contacts. One set is in the output circuit of the timer. The other set opens or closes the clutch circuit. In a case such as shown in the diagram, the clutch would probably be controlled by a solenoid. The actual details of the solenoid circuit depend on the construction of the clutch.

A different method of controlling the clutch can be utilized in the same type of mechanism as discussed above. In this case the triggering voltage also energizes a solenoid which keeps the clutch engaged as the motor turns. When the triggering voltage is released, the solenoid in the clutch deenergizes and the timing mechanism recycles itself.

Another method of releasing the clutch utilizes mechanical means. A lever can be used to release the clutch when the actuating arm applies pressure to the lever. The actual method used to release the clutch is determined by the application of the timing mechanism.

If accuracy is of prime importance when using motor timing devices, other additions must be made to the mechanisms. Most motors vary in speed when the input voltage varies. However, this relation is not true of a synchronous motor. When such a motor is used in timing devices, no regulation of voltage is necessary. But in order to make the timer accurate when other than synchronous motors are used, it is necessary to regulate either the speed or the input voltage. Circuits already exist which can regulate input voltage. If it is necessary to regulate speed in order to have still greater accuracy, a special clock mechanism is used as a governor. The torque of the motor provides the actuating force for the speed-regulating clock mechanism.



Thermal delay tube

You can easily see that the more rigorous the requirements of a timing device, the more complex that timer becomes.

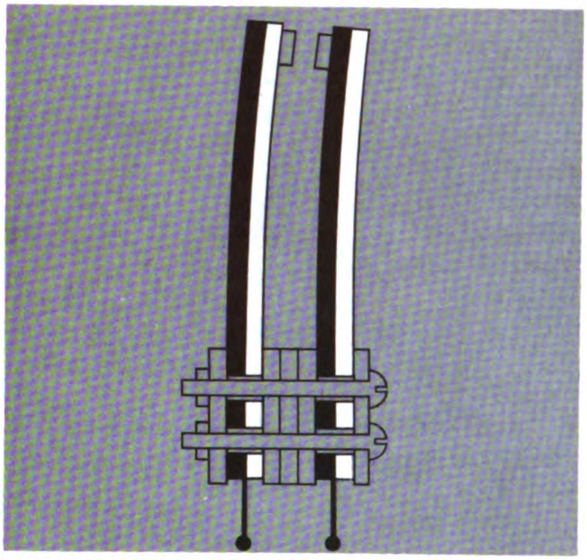
THERMOELECTRIC TIMERS. Thermal delay tubes and thermal delay relays have been used for some time to perform time delay functions. The one great advantage of timers of this type over other timing devices is the ease with which they recycle themselves. To do this, they require no additional circuitry or mechanisms. Contrasting this advantage is an important disadvantage. The disadvantage is a lack of accuracy. The accuracy of these timers is comparatively poor with respect to the other timers discussed in this section.

In the figure above, a thermal delay tube is shown. Its components are two bimetallic strips, a heating coil, a set of contacts, and a means for spacing the two bimetallic strips. A voltage is applied to the heating coil. The coil generates heat and heats up one of the bimetallic strips. This strip bends toward the other contact as temperature rises. When the bimetallic strip has heated sufficiently, the contacts close, completing the output circuit of the timing device. The amount of delay depends on the distance between the contacts; i.e., how far the bimetallic strip has to bend to close the contacts. In this case the time delay is preset and the elements are then put into a vacuum tube which prevents any further adjustment of the time delay.

Accuracy is affected if the voltage applied to the heating coil varies from its rated value. Also, if the thermal element hasn't had sufficient time to cool before the timer is used again, the accuracy is affected because the thermal element now does not require so long a time to attain the temperature at which the contacts close.

Although ambient (surrounding) temperature affects the accuracy of many thermal delay devices, it has little effect on some of the later thermal delay tubes. The ambient temperature may have a wide range, approximately −50° to −70° C, without seriously affecting the operation of the thermal device. This partial immunity to temperature change is accomplished by incorporating two bimetallic strips and fastening one of the contacts to each strip as shown in the previous illustration. When the ambient temperature changes, the temperature of each of the bimetallic strips gradually changes also. An ambient temperature change causes the same amount of change in each strip, as shown in the following illustration, because the two strips are constructed of the same two metals and are the same size.

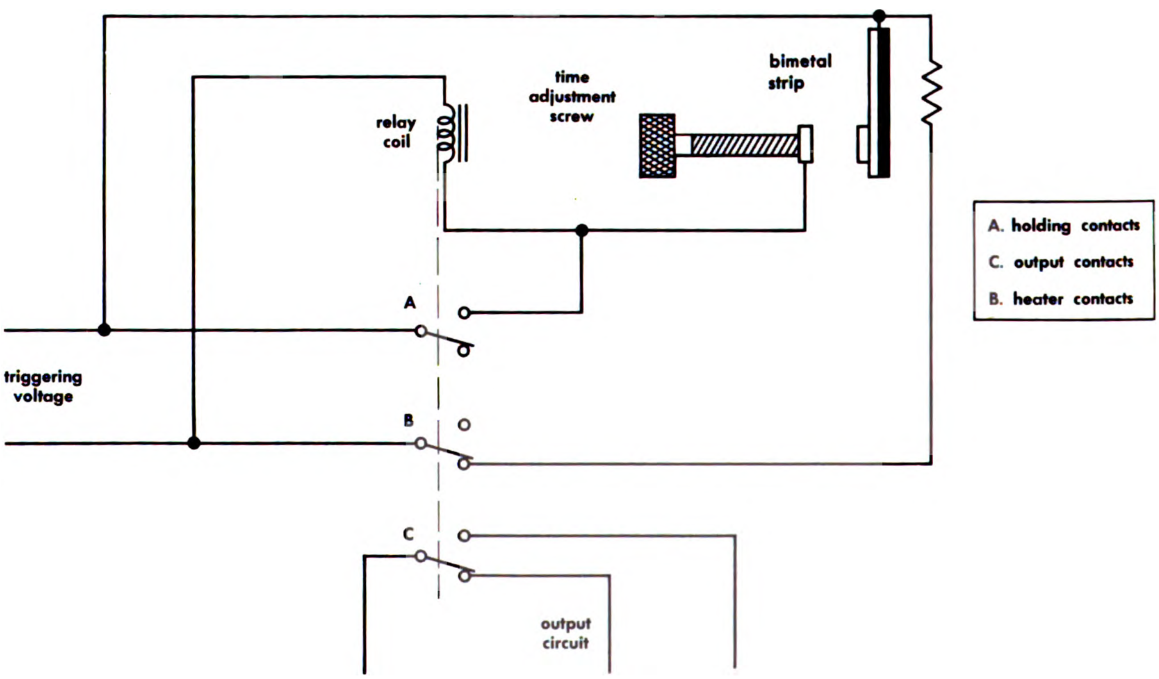
Thermal time delay relays usually do not



Thermal delay tube in extreme
ambient temperature

have this compensation for ambient temperature, so they have to be used in equipment that is temperature controlled. The time delay can be adjusted on thermal delay relays. The illustration below shows the time adjustment screw which can change the distance between the contacts. The thermal time delay relay shown can be considered a thermal device and a holding relay all built into one unit.

When the signal is applied from the triggering circuit, the voltage is impressed across the heater coils through the "B" relay contacts. As the bimetallic strip is heated, the contacts move closer together. After a period of time, the contacts close, and the voltage from the triggering circuit is applied to the relay coil. When the relay coil energizes, all of the relay contacts change. The "A" contacts parallel the contacts on the bimetallic strip and act as holding contacts, keeping the relay solenoid energized. The "B" contacts break the circuit to the heater coil, and the coil cools off. Once the relay coil has been energized and the holding contacts have closed, the bimetallic strip has served its purpose, and the timing device will not be affected if it is allowed to cool off. The "C" contacts form part of the output circuit. This type of timing device is easily recycled by taking away the triggering voltage which would deenergize the relay coil. The timer then would be ready to begin its operating cycle again.



Thermal-time-delay relay showing time adjustment

Pneumatic Timers

Pneumatic timers determine a period of time by means of a small orifice through which the air must pass.

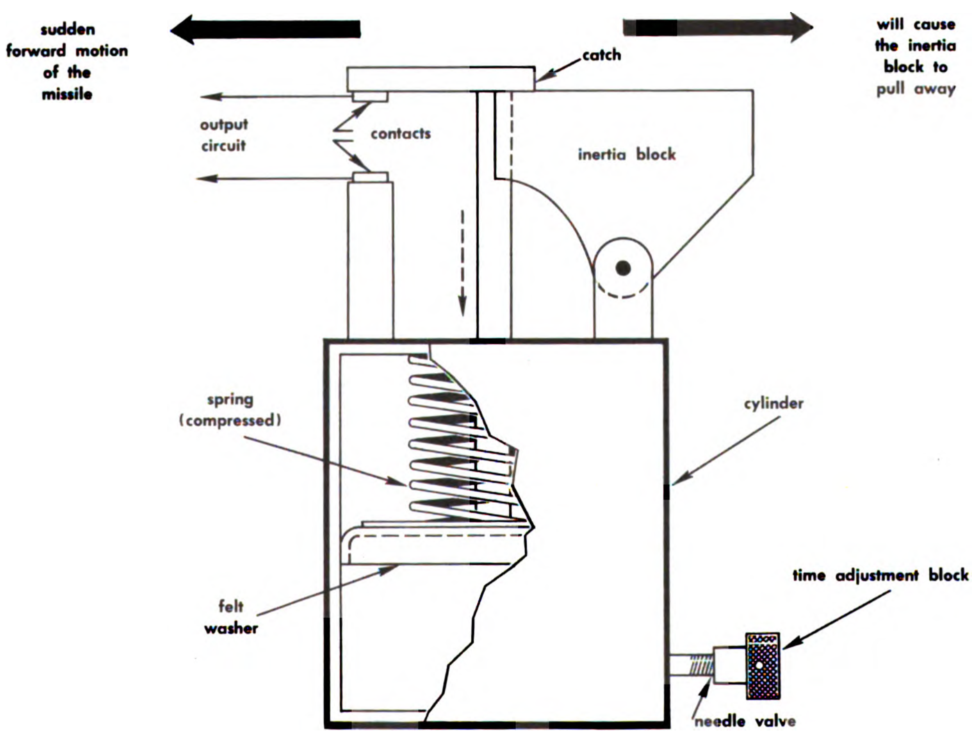
They are not very accurate in comparison with the other types that have been discussed. Factors influencing their accuracy are the effects of atmospheric pressure and temperature changes on the air stored in the timer. As mentioned, these timers get their time delay by forcing air through a small opening. If the density of air should change or if the pressure between the air stored in the timer and the atmosphere should differ, which of course would happen in a traveling missile, the accuracy would be affected. This type of timer is more suited for launch functions because it performs its function while under the same atmospheric conditions as when the time delay was adjusted.

Let's divide pneumatic timers into two general types, *piston* and *diaphragm*, for our consideration of them.

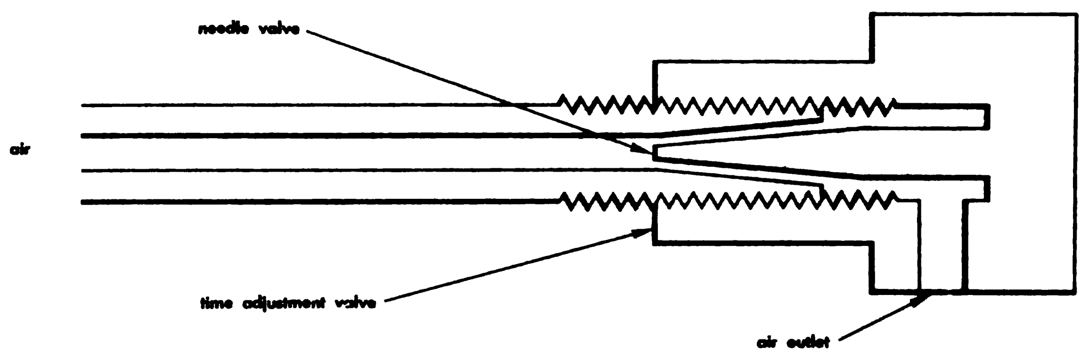
PISTON-TYPE TIMER. A piston pneumatic timer is shown below.

The action of this timer can be compared to that of a tire pump. As you pull up the handle of a tire pump, the air goes around the leather washer. The same is true of this timer. As you apply force on the handle of the tire pump, you push air out of the valve at the end of the hose. In the timer the spring applies the force to the piston, and the air in the cylinder passes through the opening in the needle valve. The spring exerts the same force each time the timer is cocked. The timer is cocked by pulling the piston up and hooking the catch over the inertia block. Because the spring exerts the same force each time, the amount of time it takes the timer to close the contacts of the output circuit depends on the size of the opening in the needle valve and the relative inside and outside air pressures. A sketch of the needle valve appears on the next page. The needle and the tube are tapered so that the gap between them will close as the adjustment screw is turned in.

The inertia block is the trigger for this timer. It consists of a block of metal which has enough mass for it to be thrown backward when subjected to large acceleration, thus the



Piston-type pneumatic timer



Needle valve of piston-type pneumatic timer

term *inertia block*. This type of triggering indicates that the timer has to be used during the launching phase.

DIAPHRAGM-TYPE TIMER. The diaphragm pneumatic timer is essentially the same in principle as the piston-type pneumatic timer. Air is taken in and then released through a small hole. The time it takes the air to escape through the hole determines the amount of time delay.

An example of a diaphragm pneumatic timer is shown in the illustration on the opposite page. The drawing shows the timer ready to perform a time delay.

The normal signal to this timer is a voltage which keeps the solenoid coil energized. When the coil is energized, the iron core is pulled down, depressing the spring. When the iron core is down, the leather diaphragm is stretched, allowing air to fill the air chamber through the air inlet holes.

The timer is triggered by taking away the voltage on the solenoid. When this is done, the spring pushes the pressure plate against the diaphragm, closing the air inlet holes and slowly pushing the air out of the chamber through the needle valve. Again, the adjustment screw controls the opening in the needle valve, thus controlling the length of time it takes the output contacts to close.

This timer is fairly accurate compared to other pneumatic timers and does not require any special mechanism to recycle itself. Reapplying the voltage to the solenoid sets up the timer for another cycle.

Timers are used in almost all control systems. It should be understood that a particular timer may be used to perform a variety of functions. The requirements of a missile, therefore, dictate the usage of a timing device. Some timers are strictly safety devices that must be used during the development of missiles.

PHYSICAL REFERENCES

There are many other references for missile control systems besides voltage and time. They can all be grouped under the heading, *Physical References*. These references are space, gravity, the earth's magnetic field, barometric pressure, and missile airframe.

In some measure they all contribute reference to most missile control systems. Primarily, they form the group of references that sensors use to establish error signals.

A brief discussion of each physical reference will be presented here, sufficient only to complete the scope of the reference unit and to familiarize you with the sensors which use these references.

Gyroscopes, because of their characteristics, are the sensors which use space for a reference. A plane of reference is established in space, and gyros sense any change from that reference.

The mass of the earth sets up a strong attraction to objects on its surface; this property is called gravity. Using this attraction as a reference, a pendulum senses the point where the attraction is greatest. Some gyros are precessed to a vertical reference

by a pendulous pickoff and an erection system. These are generally called vertical gyros, and they usually control the pitch and roll attitudes of a missile.

The lines of flux of the earth's magnetic field have served as a reference for many years through the use of the compass. A compass, as such, could not be used in a missile control system, but there is an electrical device that senses the earth's magnetic field and can be used in a control system. This instrument is known as a flux valve. Its primary use is to keep a directional gyro slaved to a given magnetic heading. This gyro usually controls the yaw channel of the autopilot.

Pressure has been used for a long time in aeronautics to indicate altitude. A missile also may use an altimeter which senses barometric pressure and produces an error signal if the missile is not at a pre-set altitude.

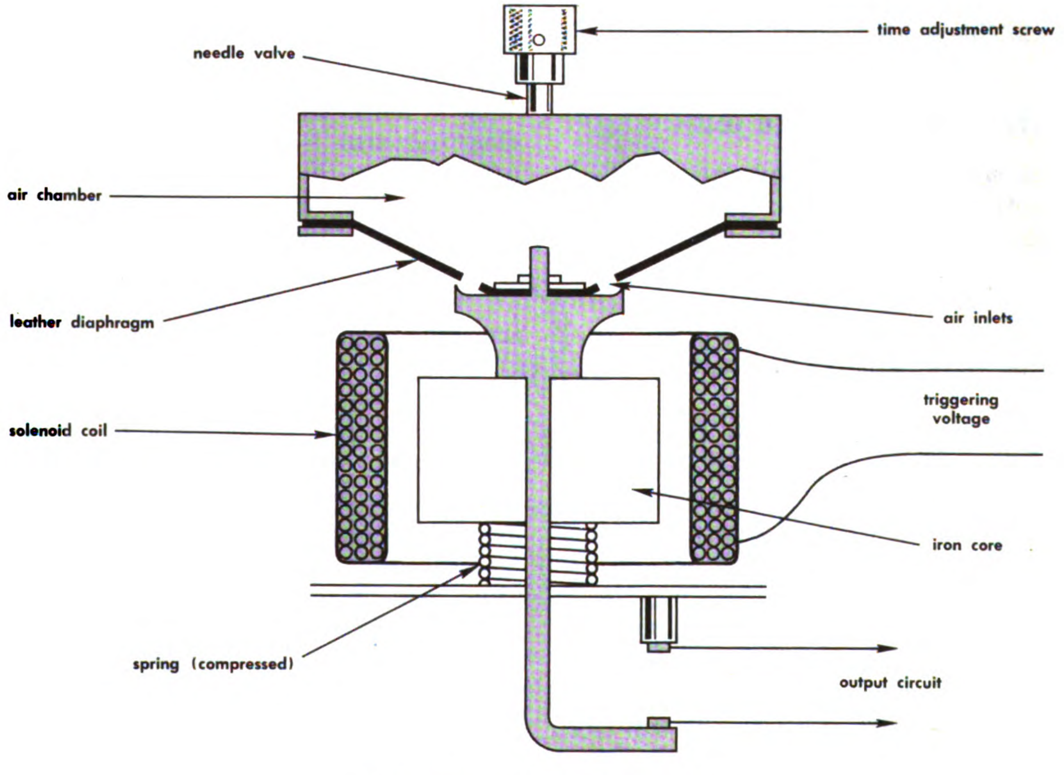
Another instrument which uses the pressure of the atmosphere as a reference is the airspeed indicator. It compares static barometric air pressure to ram air pressure. The difference derived from this comparison is indicated airspeed.

The airframe of a missile is also a reference. The displacement of the control surfaces can not be referenced to the vertical or to a given heading because the reference would have to change when the missile attitude changes. It is easier to give control-surface displacement with respect (or with reference) to the missile airframe.

Selsyns are used to indicate the angular position of the control surfaces with respect to a missile airframe. Also, potentiometers may be used to indicate angular position when the potentiometer is fastened to the missile airframe and its wiper arm is moved by the control surface.

The autopilot sensors have been classified here according to the reference they sense. For further information about sensors, refer back to Section 1 of this chapter.

This concludes the comments on reference units. In the next chapter, amplifier units are covered.



Diaphragm-type pneumatic timer

Amplifier Units of Control Systems

An amplifier is a device, usually containing one or more vacuum tubes, whose output is an enlarged reproduction of the input signal. Voltage amplifier units are designed to develop the greatest amplified voltage possible across the load in the plate circuit of the amplifier. Power amplifier units are designed to deliver large amounts of power to a plate circuit load without regard to voltage.

Both *voltage* and *power* amplifiers are used in missile control systems. Some of the less conventional amplifiers are included in the discussion in this section.

VOLTAGE-TYPE AMPLIFIER UNITS

Electrical signals in a control system may be either AC or DC. Since some system components may require DC, and other components in the same loop may require AC, it is desirable to have devices which can convert electrical signals from DC to AC and vice versa. Modulators perform this function.

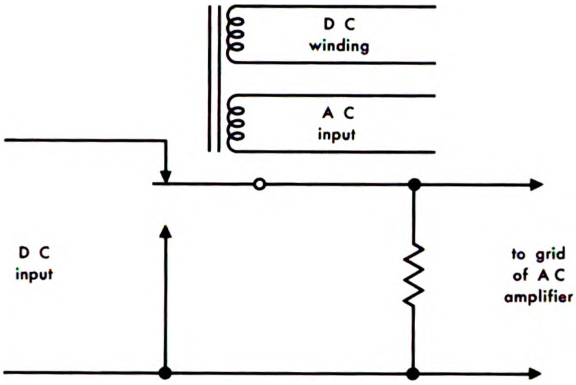
A modulator is a device which alters the amplitude, frequency, or phase of a wave (carrier) in conformity with the variations of an input signal. In a control system, modulators generally are used to convert polarized DC signals to properly phased AC signals. There are several types of modulators which will perform this function in control systems. We will consider the mechanical and vacuum-tube choppers and the rectifier modulator. Commutators and paraphase amplifiers also are covered in this part on voltage amplifiers.

Mechanical Choppers

A mechanical chopper is simply a synchronized switch. Its output line is connected alternately to its input line and to ground, or zero-level DC. The explanation which follows refers to the illustration to the right above.

The coil is energized by AC of the appropriate frequency. In order to prevent the vibrating arm from being attracted twice by the coil during each cycle, some direct current is passed through the coil, or through another coil wound on the same core. Thus, when the AC flows in one direction, its magnetizing force is aided by that of the DC, and the vibrating arm is attracted. But when the AC flows in the other direction, its magnetizing force is cancelled by that of the DC, and the attraction on the arm is reduced. The arm thus vibrates at the frequency of the AC. To reduce the power required to drive the arm, the arm usually is tuned so that it vibrates naturally at the proper frequency. The output is then a square wave, having the level of the DC input for a half-cycle and having zero level for the other.

Blocking condensers in the amplifier that follows the chopper remove the DC component, so that the signal becomes a square wave with an amplitude proportional to the DC signal level and a phase which reverses when the DC signal changes polarity. This amplifier also discriminates against high-frequency signals, so that the square wave ultimately is converted to a sine wave at the frequency of the chopper switching.

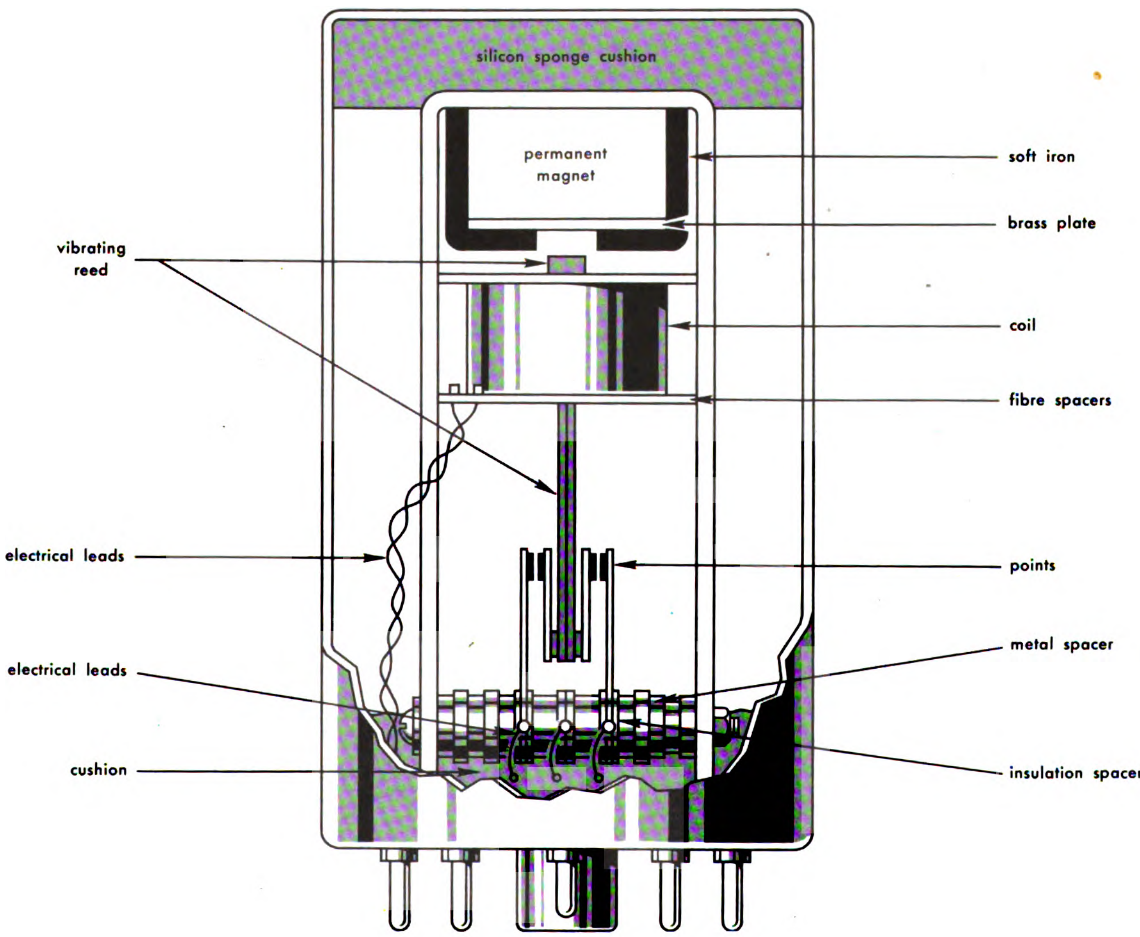


Mechanical chopper

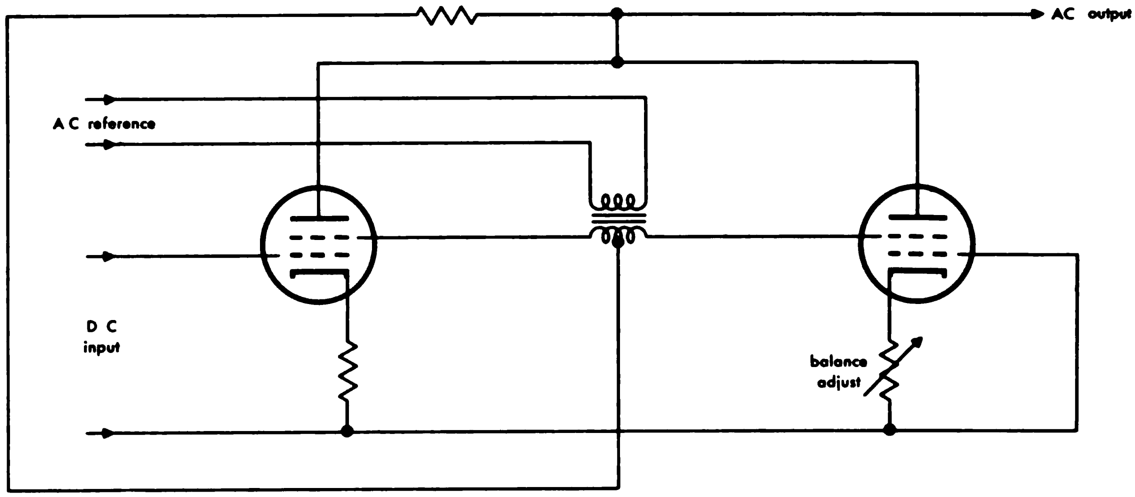
A cutaway view of a mechanical chopper is shown in the next illustration. In this particular chopper a permanent magnet is used in place of the DC winding shown in the above chopper.

Vacuum-Tube Choppers

A vacuum-tube chopper amplifier circuit, shown on the following page, consists of two tubes, one having a pulsed DC signal on its control grid and the other a fixed voltage. Alternating current of the proper frequency is applied to the screen grids, the voltages at the two screen grids being 180 degrees out of phase with each other. When the tubes carry the same current, the effect of raising the voltage on one screen is exactly balanced by that of lowering the voltage on the other screen; consequently, the total current flowing in the common load resistor is unchanged, and there is no AC output. This condition is no longer true in the presence of a DC signal. AC appears across the load resistor with an amplitude proportional to the DC level and with a phase which reverses when the DC polarity is changed.



Cutaway view of mechanical chopper



Vacuum tube chopper circuit

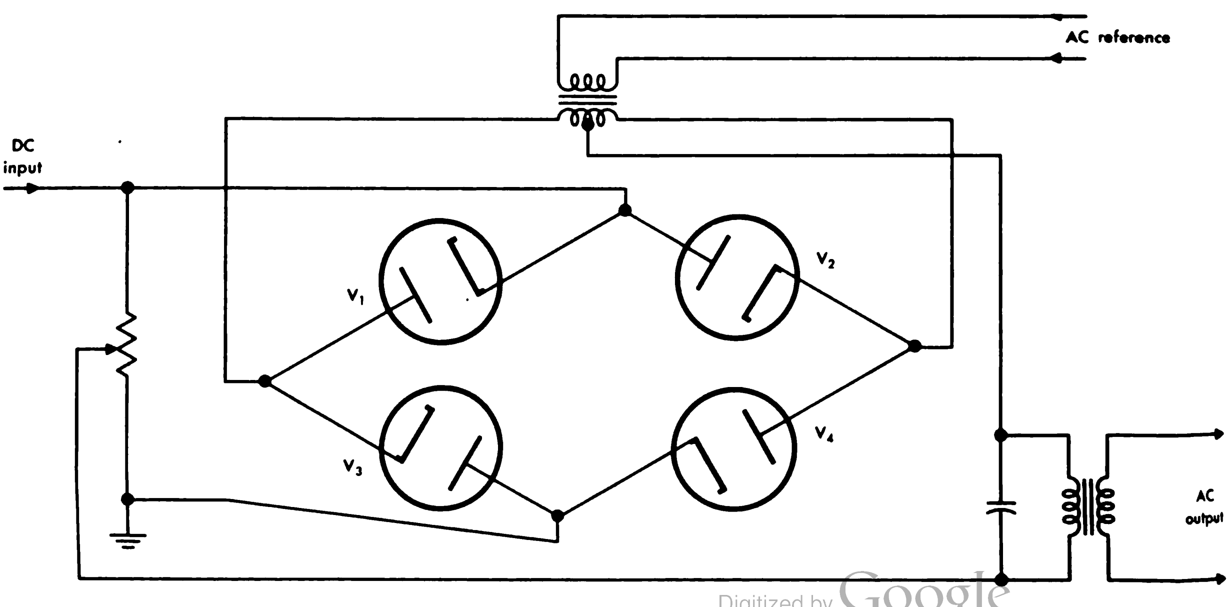
The variable cathode resistor in one tube is used to balance the currents so that the output drops to zero when the input is zero.

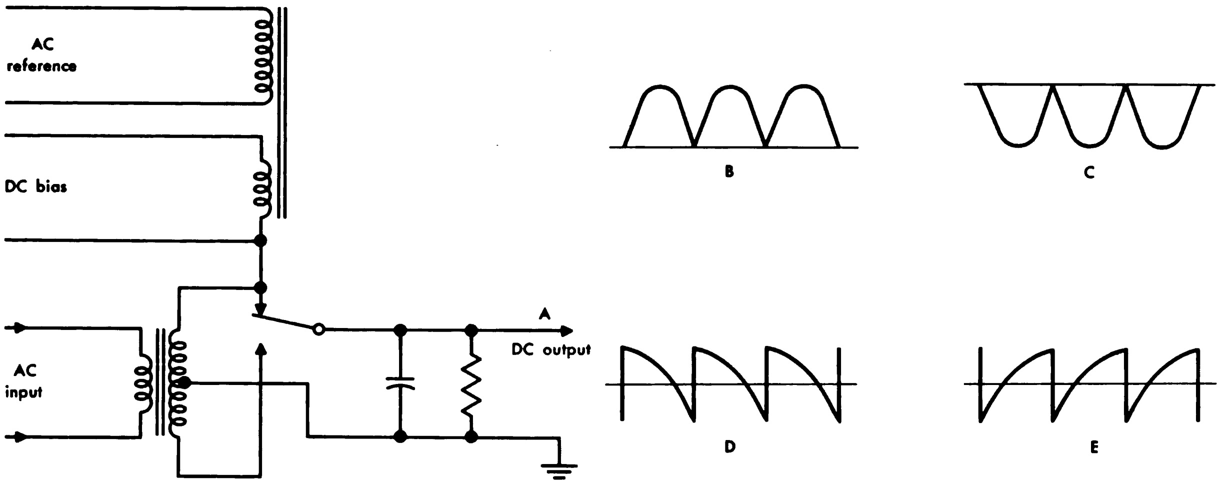
Vacuum-tubes can be employed in many other ways to produce the same result, as does the circuit just described.

Rectifier Modulators

A rectifier modulator, as illustrated below, is another type of synchronous switch. During a half-cycle, an AC reference signal causes current to flow through "V1" and "V2;" during the other half-cycle, it flows through "V3" and "V4." The nonconducting diodes appear as open circuits, and the conducting diodes appear as relatively low resistances. One end of the input resistance is, therefore, connected to the center tap of a divider across the output of the reference AC transformer. It is at the same voltage as the center tap on the transformer secondary. The output voltage, then, is equal to that appearing across one-half of the input resistor. Because of the switching action, the output circuit is subjected alternately to the voltages across the upper and the lower halves of the input resistor, both voltages being measured with respect to the center of the resistor. The resulting output voltage is equal to half of the input voltage during half of the AC cycle, and is equal to half of the amount of the input voltage but opposite in polarity during the other half of the AC cycle. The action is the same as that of a mechanical chopper. Since one side of the input resistor is usually grounded, it is necessary to take the output through an isolating transformer in order to remove the DC component.

Rectifier modulator circuit



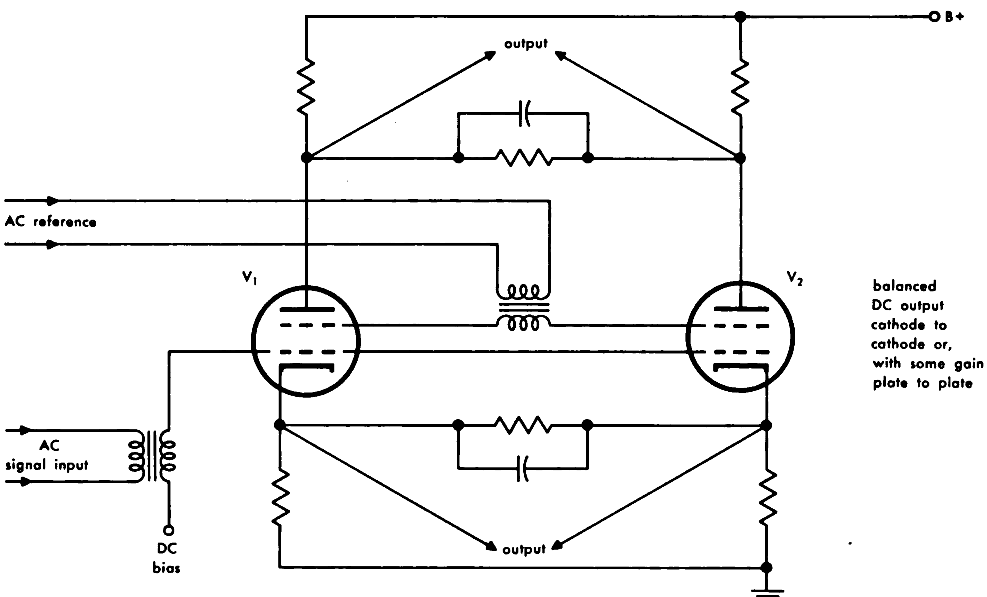


Synchronous switch circuit and output wave forms

Commutators

A commutator, sometimes known as a phase-sensitive detector, makes the same conversion as a modulator, but in the reverse order. It receives an AC input and develops a DC output. For an input of one phase, it develops a positive DC output. For an input with a phase difference of 180 degrees, it develops a negative output. For inputs differing from these by 90 degrees, it develops no DC output. There are several types of commutators, three of which are: the synchronous mechanical switch, the vacuum-tube commutator, and the rectifier commutator.

SYNCHRONOUS MECHANICAL SWITCH. A synchronous mechanical switch, such as shown in the illustration above, depends on a vibrating arm similar to that of a mechanical chopper. Assume that the voltage at the upper contact is positive during the half-cycle when the arm is up. During the next half-cycle, when the arm is down, the volt-



Vacuum-tube commutator circuit

age at the lower contact is also positive. The resulting output is similar to that of a full-wave rectifier. After being filtered, the output has a DC component proportional to the amplitude of the AC. If the phase of the input is reversed, the output is negative DC.

The signals in these two examples are shown by wave forms "B" and "C" in the illustration. For signals differing from these by 90 degrees, the voltage at each contact reverses during the time the arm makes connection, and the average voltage is zero as shown by wave forms "D" and "E." There is, therefore, no DC component and no output from the circuit after filtering.

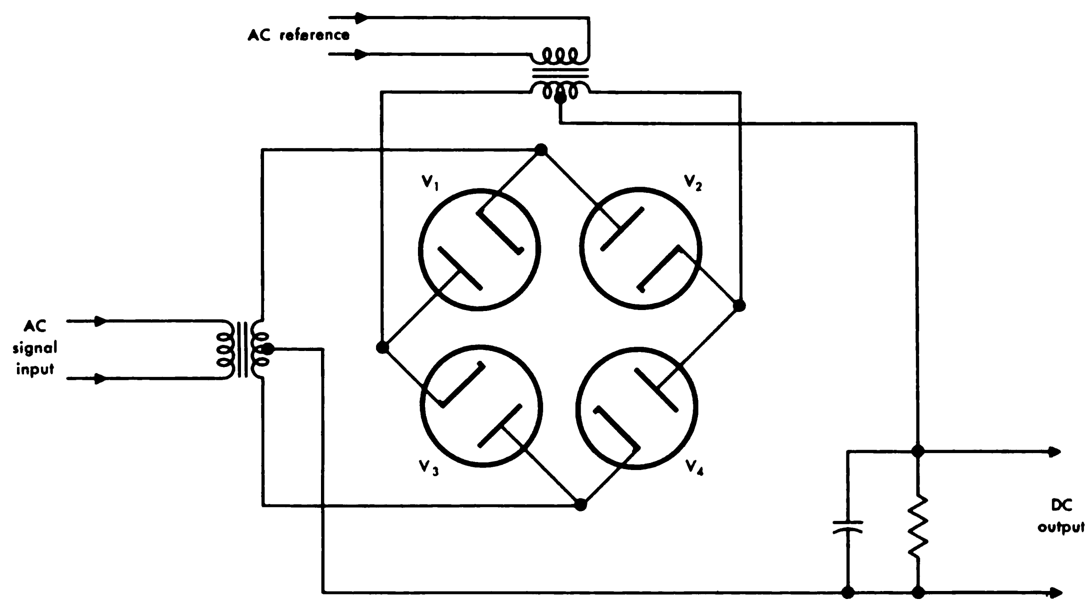
VACUUM-TUBE COMMUTATOR. A number of vacuum-tube circuits can serve as commutators. One such circuit is shown on the preceding page. The DC bias on the control grids is sufficient to hold both tubes at cut-off. Assume that the voltage on the control grids moves in the positive direction at the same time as that on the screen grid of tube "V1." During the half-cycle of current flow, the screen grid of "V1" always is more positive than the screen of "V2," and "V1" draws more current. The cathode of "V1," therefore, is more positive than the cathode of "V2," and the plate of "V1" is more negative than the plate of "V2."

A phase reversal of the control grid signal causes "V2" to draw more current and reverses the polarity of both the cathode-to-cathode and the plate-to-plate voltages. If the grid signal differs by 90 degrees from the screen signal, "V1" draws more current than "V2" during one-half of the conducting half-cycle. The situation is reversed during the remainder of the conducting period. The DC component of the output is therefore zero. A circuit of this sort is useful where DC output need not be developed with respect to ground.

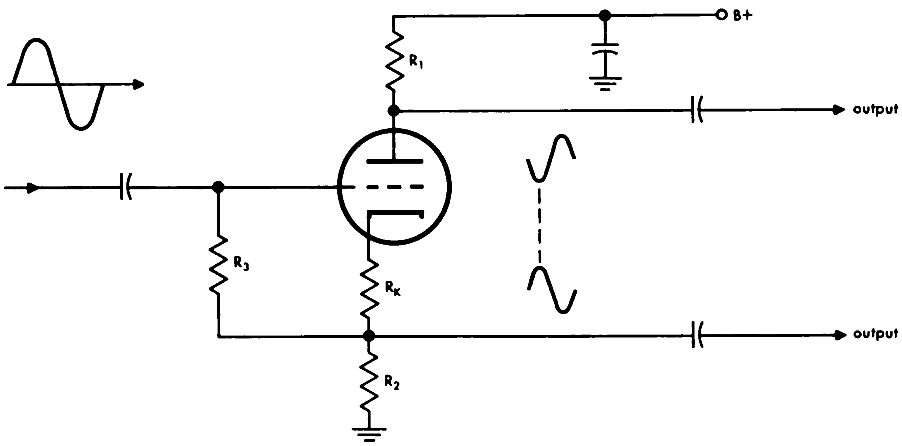
RECTIFIER COMMUTATOR. A rectifier commutator, as illustrated below, acts in much the same way as the rectifier modulator described previously. During a half-cycle, the reference AC causes "V1" and "V2" to conduct; during the other half-cycle, it causes "V3" and "V4" to conduct. The output circuit is thus connected alternately from the center tap of the input transformer to the junction of the upper diodes and from the center tap to the junction of the lower diodes.

Paraphase Amplifiers

Paraphase amplifiers are sometimes used in place of transformers to operate push-pull circuits. A paraphase amplifier is a combination amplifier and phase inverter.



Rectifier commutator circuit



Paraphase amplifier

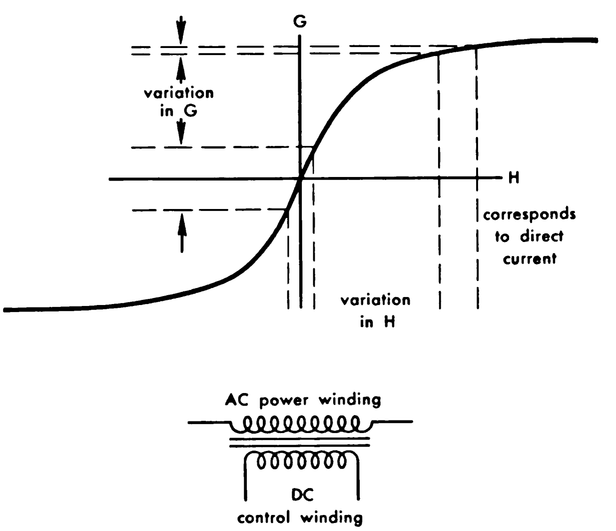
The illustration above is a circuit diagram of one type of paraphase amplifier. In this circuit the outputs are taken across the resistors "R1" and "R2." These resistors are of equal value, and, since the same current flows through both, equal voltages are developed across them. The voltages across these resistors are opposite in polarity since the output is taken from the negative end of "R1" and positive end of "R2."

POWER-TYPE AMPLIFIER UNITS

As stated earlier, power amplifiers deliver large amounts of power to a plate circuit load. Before taking up power amplifiers, let's consider saturable reactors which are used in power amplifier units of missile control systems.

Saturable Reactor

A saturable reactor is, in effect, an inductor whose inductance can be varied by means of a control current. A magnetizing force (H) is developed when current flows through the winding of an iron-core inductor. This, in turn, sets up a magnetic field (G) in the core. When the current is changed, the accompanying change in "H" causes a change in "G." The change in "G" induces voltage in all of the windings that encircle the iron core. The induced voltage is, of course, proportional to the rate at which "G" changes. If "G" is proportional to "H," the induced voltage is proportional to the rate at which the current changes. In some types of core material, the relation between "G" and "H" is like that shown in the illustration below. A winding on such a core shows induced voltage when AC is passed through it. Now, if DC is added in the same winding or in another winding, the variation in "H" no longer takes place about an average value of zero. Instead, "H" oscillates about a value which corresponds to the DC current. If this point is sufficiently far beyond the bend in the "G-H" curve, there will be only a small change in "G" and a small induced voltage. The result is that an inductor made of such material can be designed to have a high inductance when no DC flows in the windings, and a much lower inductance when DC does flow.

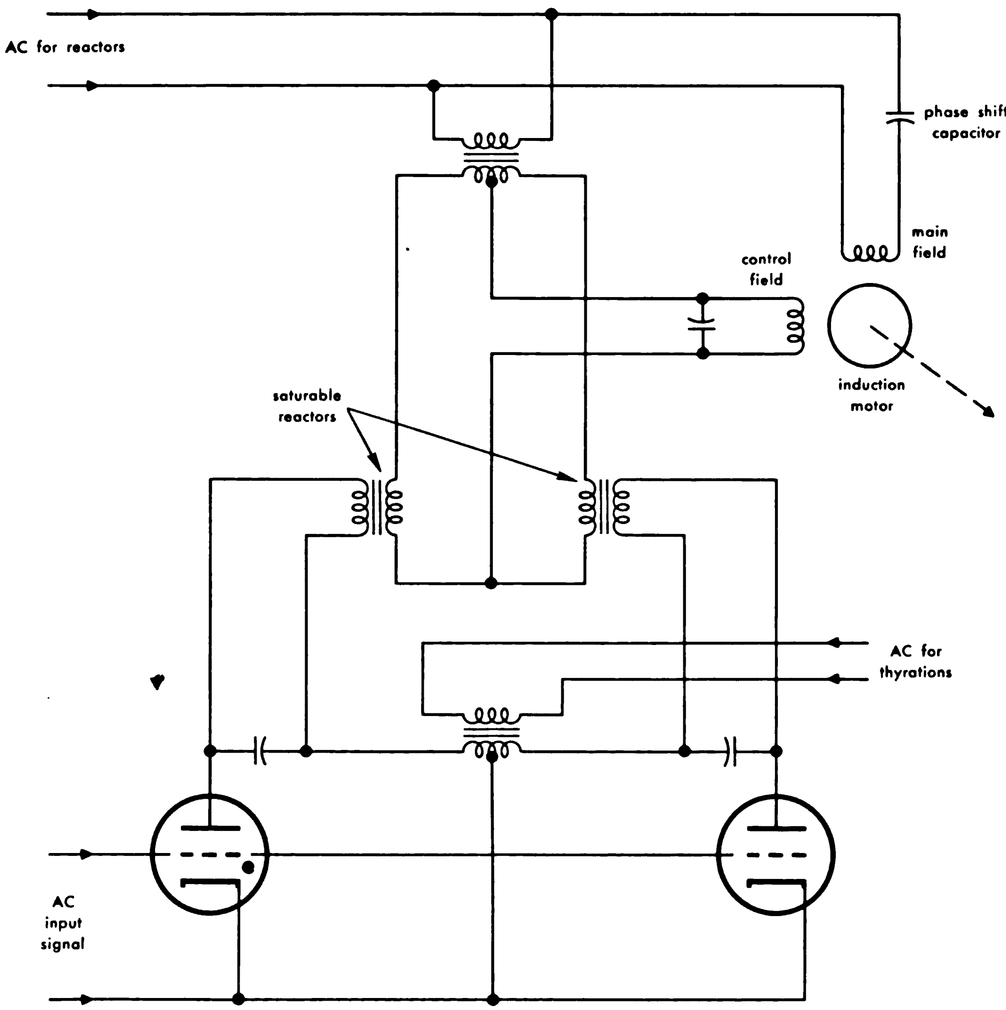


G-H curve for saturable reactor

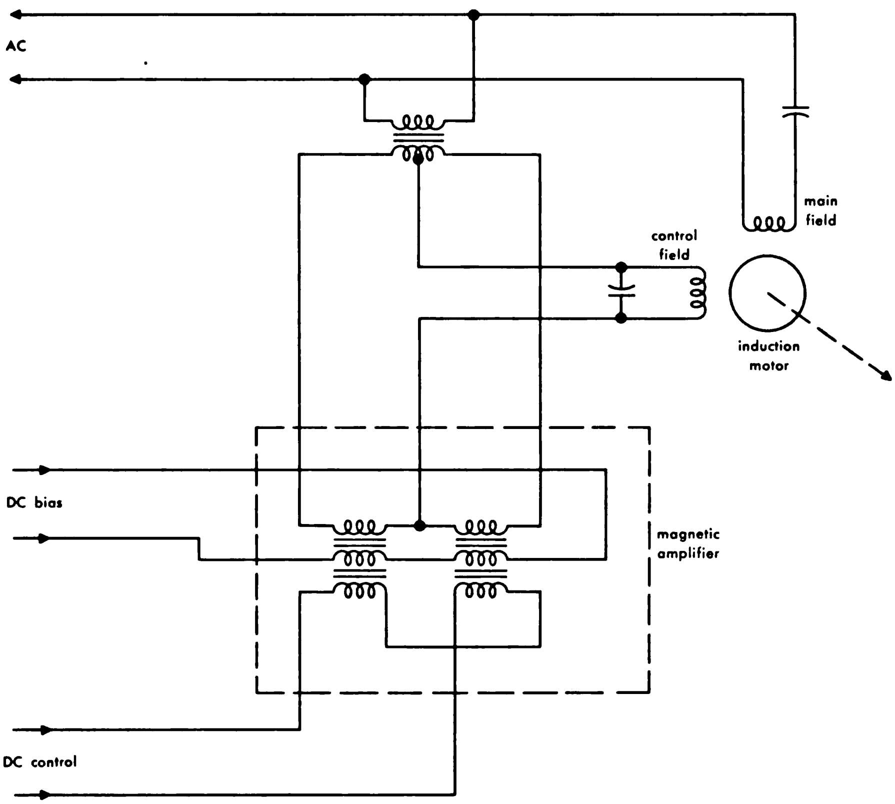
Saturable-Reactor Control of AC Inductance Motor

The illustration below shows a pair of saturable reactors connected so as to control an AC induction motor. The main field of the motor is connected to the line in the usual way. One end of the control field winding is connected to the center tap of a transformer fed by the line. The other end is connected through a pair of saturable reactors to both ends of the transformer. When no DC flows through either reactor, the control field is not energized. When DC flows through either reactor, the reactor's inductance is decreased so that the control field is connected to one end of the transformer through a high impedance and to the other end through a low impedance. The control field is, therefore, energized in one direction or the other, and forward or reverse torque is developed by the motor. The DC through the reactors is supplied by a pair of thyratrons. These tubes receive an AC error voltage and are connected to an AC plate supply.

Note that the AC for the motor and the reactors may be of one frequency, and the error voltage and the thyratron anode supply may be of another. Vacuum tubes can be used instead of thyratrons if the error signal is AC. If the error signal is DC, the signal might possibly be used directly to saturate the reactors, or a DC amplifier may be used when the signal is not strong enough to produce the desired degree of saturation.



Saturable-reactor control circuit for an AC motor



Magnetic amplifier control circuit for an AC motor

Magnetic Amplifiers

A magnetic amplifier uses saturable reactors. These reactors have extra windings that carry a direct current called the bias current. The diagram above shows a magnetic amplifier circuit controlling an AC motor. It is nearly the same as the circuit for saturable reactor control except that the two reactors have their control windings connected in series. The direction in which the bias current flows is so chosen that the magnetizing force developed by the bias current aids the force developed by the control current in one reactor and bucks the force in the other. In the absence of a control signal, the DC's are the same in each reactor, and the control field of the motor is not energized. The control current and the bias current are, in effect, added in one reactor and subtracted in the other. Thus, both reactors are partially saturated in the beginning, and the control current saturates one even more and reduces the saturation of the other. A DC control signal is required, but this can be obtained from an AC error signal by means of a phase-sensitive detector.

Both the magnetic amplifier and the saturable reactor suffer from the fact that, because of the inductance of the control winding, a perceptible time is required to establish the control current. A magnetic amplifier is more desirable than a saturable reactor because it has neither moving parts nor vacuum tubes.

This duscussion of amplifier units is not, of course, a complete discussion of all types of amplifiers used in control systems. The intent here has been to familiarize you with the less conventional types.

Controller Units of Control Systems

A controller unit in a missile control system controls the operation of the actuator which is responding to an error signal it receives from the sensing element. In some systems, an amplifier, whose output is being applied to a motor, is the controller unit. However, in this section we will consider only those controller units, other than amplifiers, which control actuator units in a system. Note that the "controller" block follows the "amplifier" block in the diagram below.

A description of solenoids follows because they are important components of controller units.

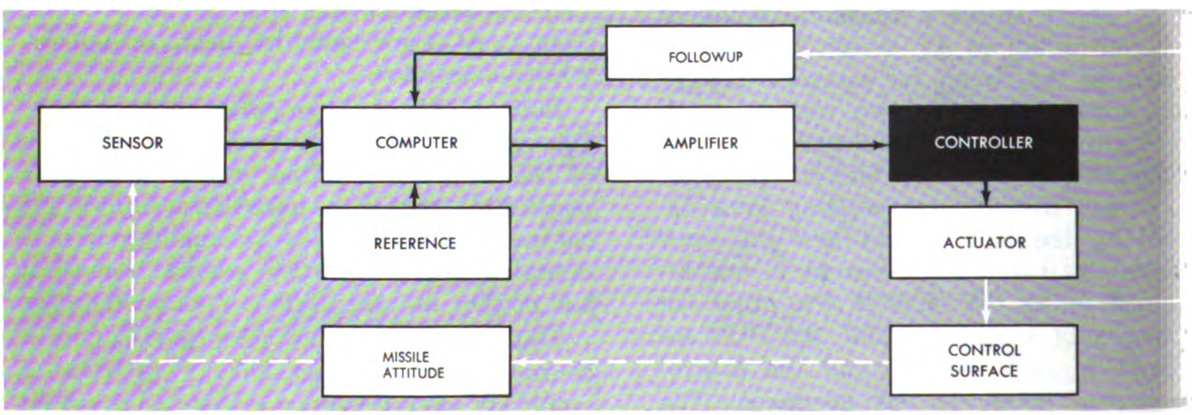
SOLENOIDS OPERATE VALVES AND RELAYS

A solenoid consists of a coil of wire wound around a hollow cylinder. It is used to produce a magnetic field. If a movable core of soft iron is placed inside the cylinder, the field of the coil tends to center the core into the coil when current is flowing. Solenoid coils with movable cores are used for remote control of various units such as solenoid-operated valves and relays.

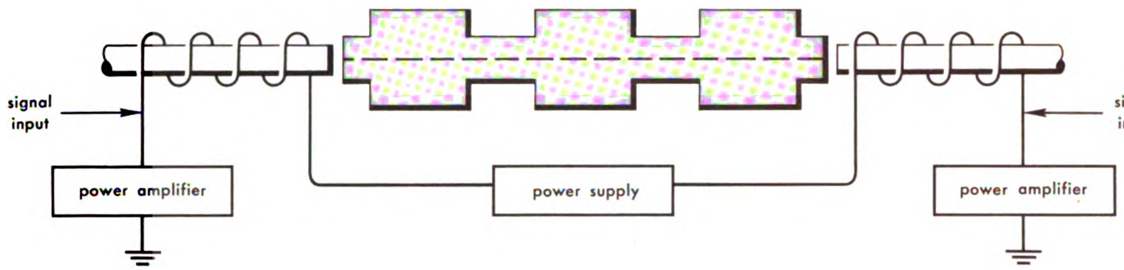
If two solenoids are arranged as shown in the illustration above right, they can control movement of a needle valve in a hydraulic system, pneumatic system, etc.

TRANSFER VALVE CONTROLS FLUID TO ACTUATING DEVICE

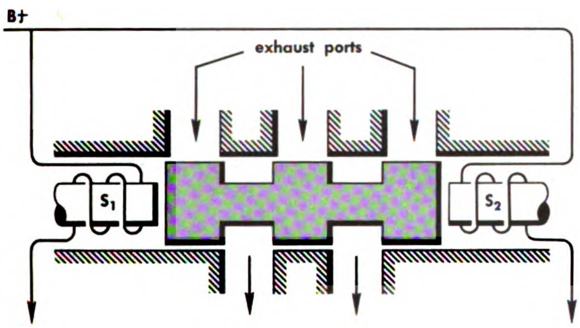
Transfer valves are used in control systems having hydraulic actuators. The transfer valve unit illustrated on the right consists of a double-acting solenoid which controls the position of a spool in the transfer valve unit. The position of the spool determines



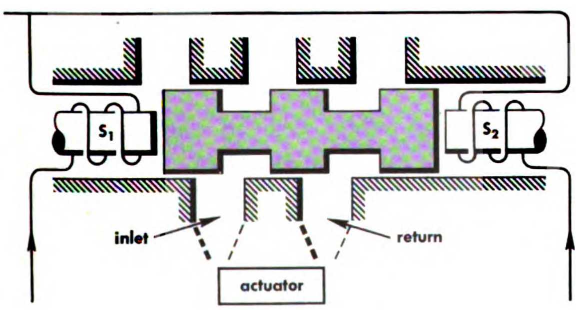
Basic missile control block diagram



Double acting solenoid



Transfer valve (closed)



Transfer valve (open)

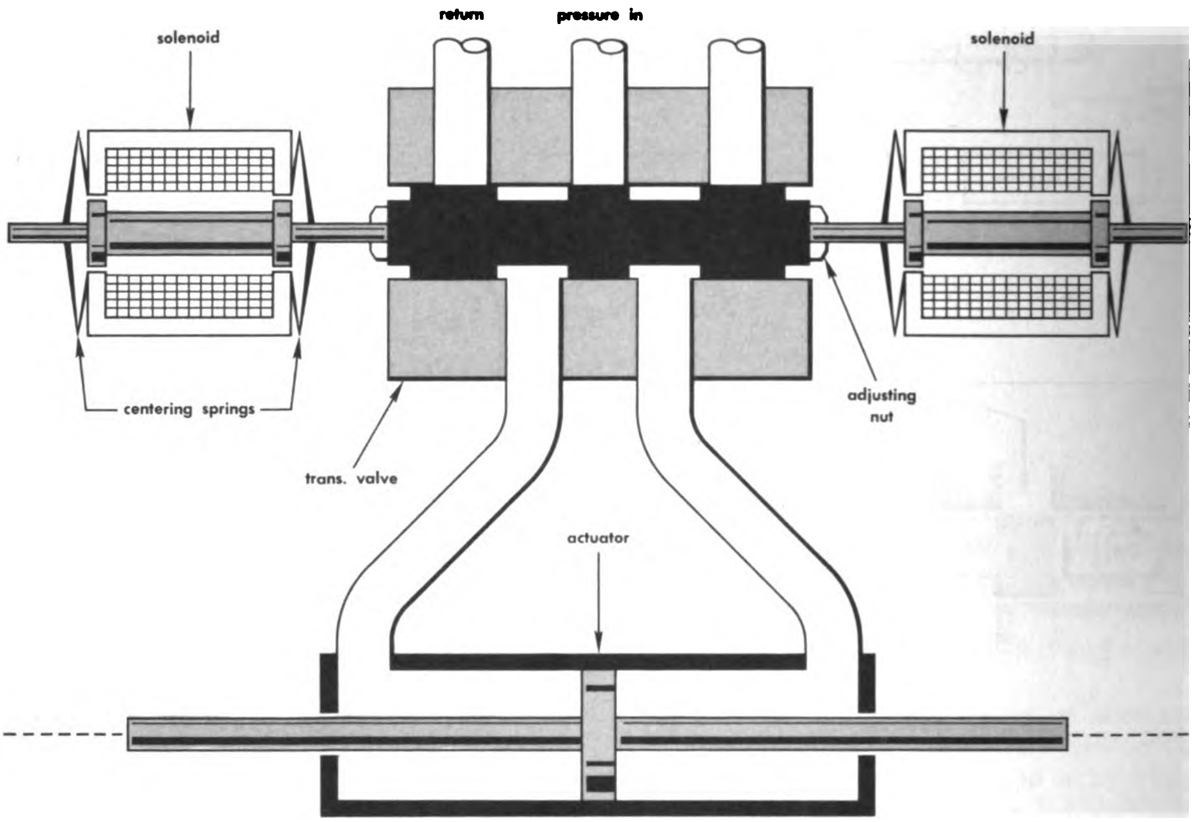
how much fluid is allowed to flow to the hydraulic actuating device. In the figure directly above, the spool valve is shown displaced from its center position to allow fluid to flow under pressure to the actuating unit. Solenoid "S1" has a heavier current flow through its winding than solenoid "S2;" thus "S1" attracts the spool. This action opens the ports. Now the hydraulic fluid is able to flow through the transfer valve to one side of the actuator. The fluid on the opposite side of the actuator is returned back through the exhaust port as shown. A more complete drawing of the transfer valve and actuator is on the following page.

RELAY SWITCHES CONTROL HEAVY-CURRENT CIRCUITS

Relay switches are used for remote control of heavy-current circuits. They are placed directly between the source of power and the controlled unit so that the cables carrying heavy current will be as short as possible. A relay switch consists of a coil or solenoid, an iron core, and fixed and movable contacts. Small wires connect the solenoid coil terminals with the source of power which is the control signal. When a control signal is present, an electromagnetic field is set up around the coil.

In the relay switch shown first on page 243, the iron core is fixed. When the control signal is present, the core is magnetized by the field set up around the coil. The pull of the core on the piece of soft iron overcomes the force of a spring, thus closing the contacts. This action completes the heavy-current circuit. When the control signal is removed, the field around the coil collapses. Then the spring separates the contacts, breaking the heavy-current circuit.

In another type of relay switch, part of the core is movable, as indicated in the right hand drawing on page 243. Contacts are attached to the coil mounting but insulated from it. When the control switch is closed, the field around the coil causes the movable parts of the core to be drawn into the coil, closing the contacts and completing the heavy-current circuit. When the control switch is opened, the field around the coil collapses, and the return spring returns the movable core to its original position, separating the contacts.



Hydraulic transfer valve and actuator

The more quickly a circuit carrying a large current is opened, the less it will arc and the less the switch contacts will be burned. Relay switches used to control the circuits of large motors have strong return springs which open the switches quickly.

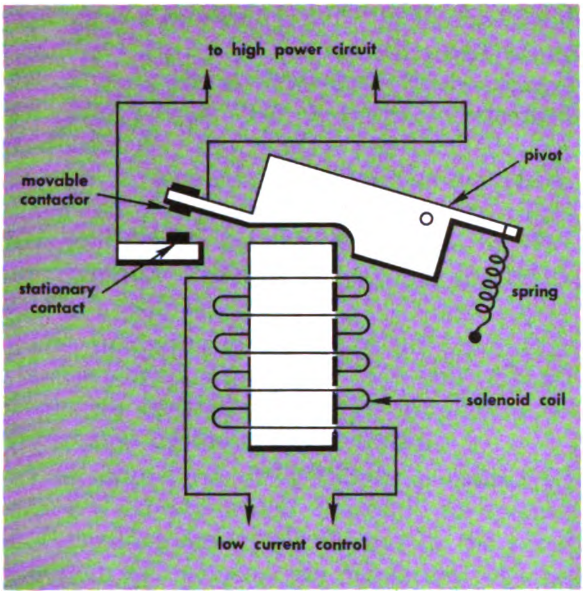
Relay switches have either an insulating spacer on each coil terminal or an insulating spacer on one coil terminal and a metal spacer on the other. If a metal spacer is used, it grounds the terminal to the coil case. Thus, no ground wire to that terminal is required.

To permit the circuits controlled by heavy-duty relay switches to carry heavy currents and still protect them against short circuits, a special type of fuse called a *current limiter* is used. A current limiter permits the large overloads required for starting motors but "blows" before the circuit is damaged if the overload is continued.

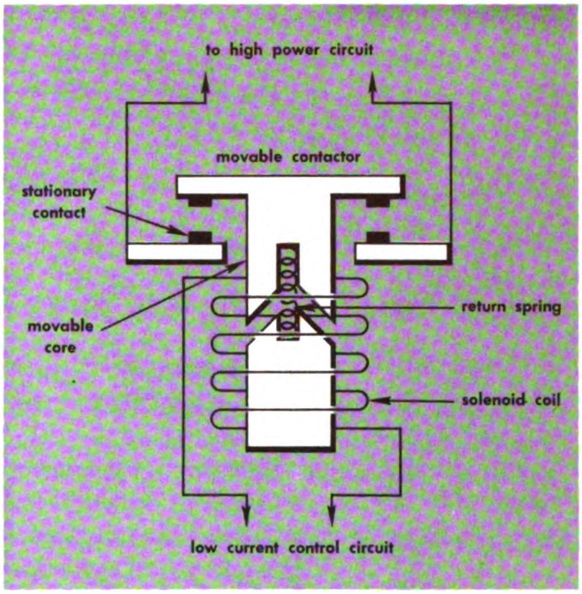
In the illustration opposite, two air-pressure input lines of an air-actuated relay are connected to the relay, one to each side of a diaphragm. This type of relay is used in a pneumatic-electric control system. When an error is present in the system, an air pressure is applied to one side or the other of the diaphragm. The diaphragm is forced over to complete an electrical circuit connection from terminal "C" to one of the other two terminals, "A" or "B." In this manner power is transferred across the relay into a suitable actuating circuit. Similar relays might be found in a hydraulic-electric system in which an electrical circuit is actuated by hydraulic pressure rather than air pressure.

AMPLIDYNES ARE A SPECIAL TYPE OF DC GENERATOR

An amplidyne consists of a DC generator rotated by an external motor. The DC generator can be considered an amplifier since a small amount of power applied to the field coil controls many times as much power in the

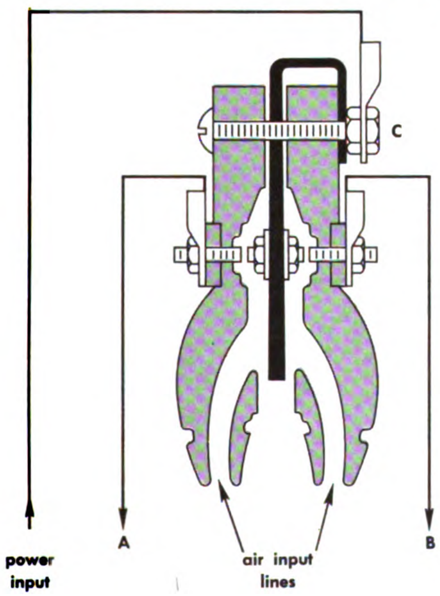


Relay switch with fixed iron core



Relay switch with part of iron core movable

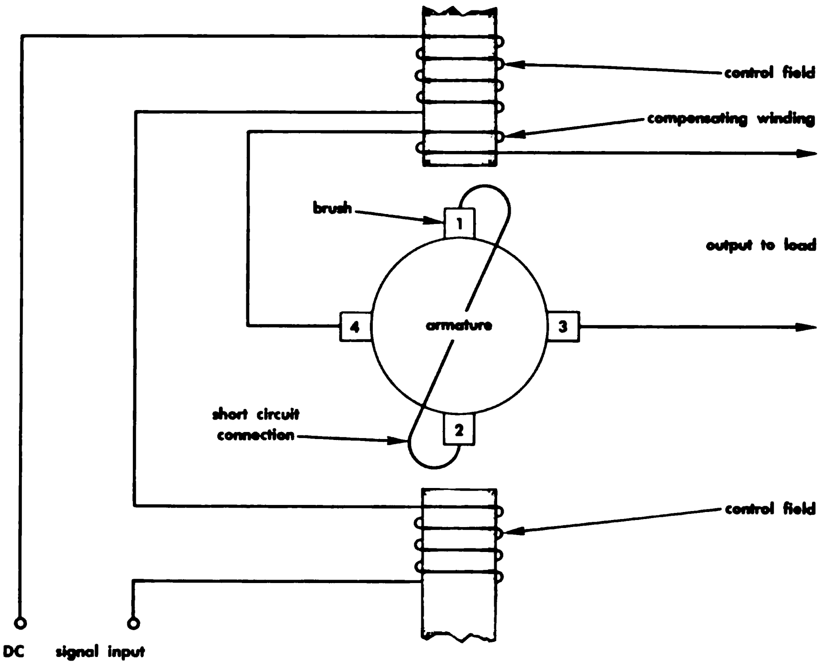
output of the generator. An amplidyne schematic is shown on page 244. The DC input control voltage is used to excite the field structure which is shown as a pair of poles above and below the armature. Rotation of the armature in the control field induces a voltage in the coils as they pass through this magnetic field. Since these coils are in a vertical plane when crossing the control field, the voltage appears at the upper and lower brushes numbered 1 and 2. If a load were connected to these two brushes, the device would be an ordinary DC generator. However, note that in the drawing, brushes 1 and 2 are tied together, or short-circuited. The resulting high current flowing in the armature produces a second magnetic field which is much stronger than the control field and at right angles to it. The lines of force of this second magnetic field lie in a horizontal plane. Armature coils cutting this second magnetic field have a voltage induced in them which appears across the second set of brushes, 3 and 4, located on the horizontal axis. This induced voltage is the output of the amplidyne.



Air-actuated relay

A problem of major concern relative to the amplidyne is the cancellation of the control field. When load current flows, it flows through armature coils which produce a magnetic field. The magnetic field tends to cancel the control field. This cancellation is an undesirable feature because it reduces the amplification and causes poor regulation of the amplidyne system.

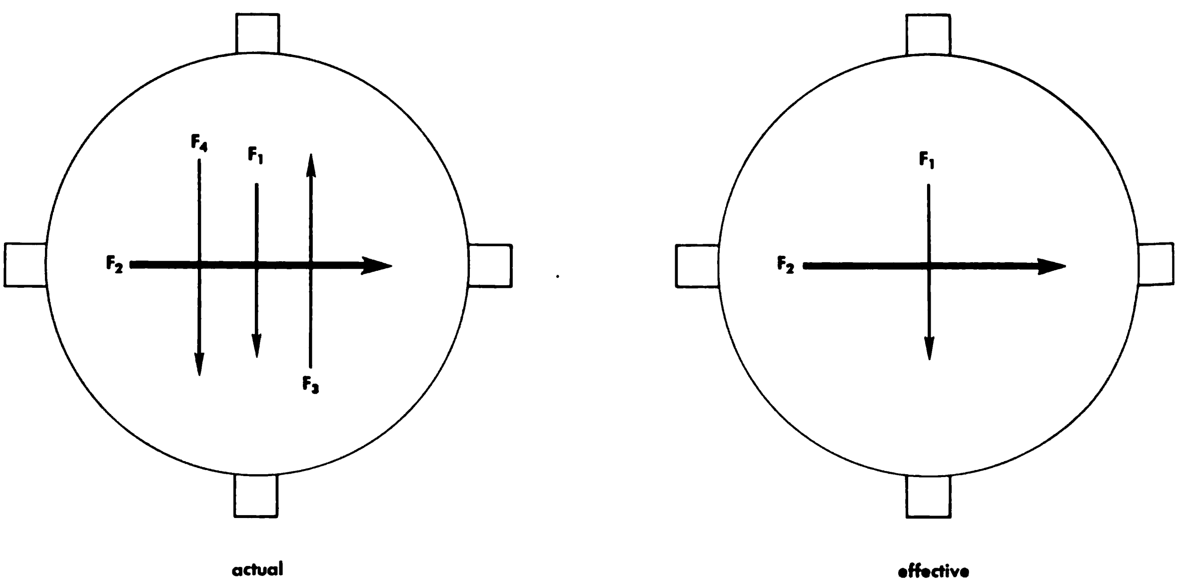
It is possible to correct this cancellation tendency by adding auxiliary windings to the field structure and passing the output current



Basic schematic of an amplidyne

through them. The magnetizing force produced by passing the output current through the auxiliary field can be made to balance that produced by the flow of output current through the armature coils. The various magnetic fields are represented by arrows in the following diagram. As shown, magnetic fields in an amplidyne occupy certain positions and have certain directions, with respect to each other.

"F_1" represents the control field produced by the input voltage. "F_2" is the magnetic



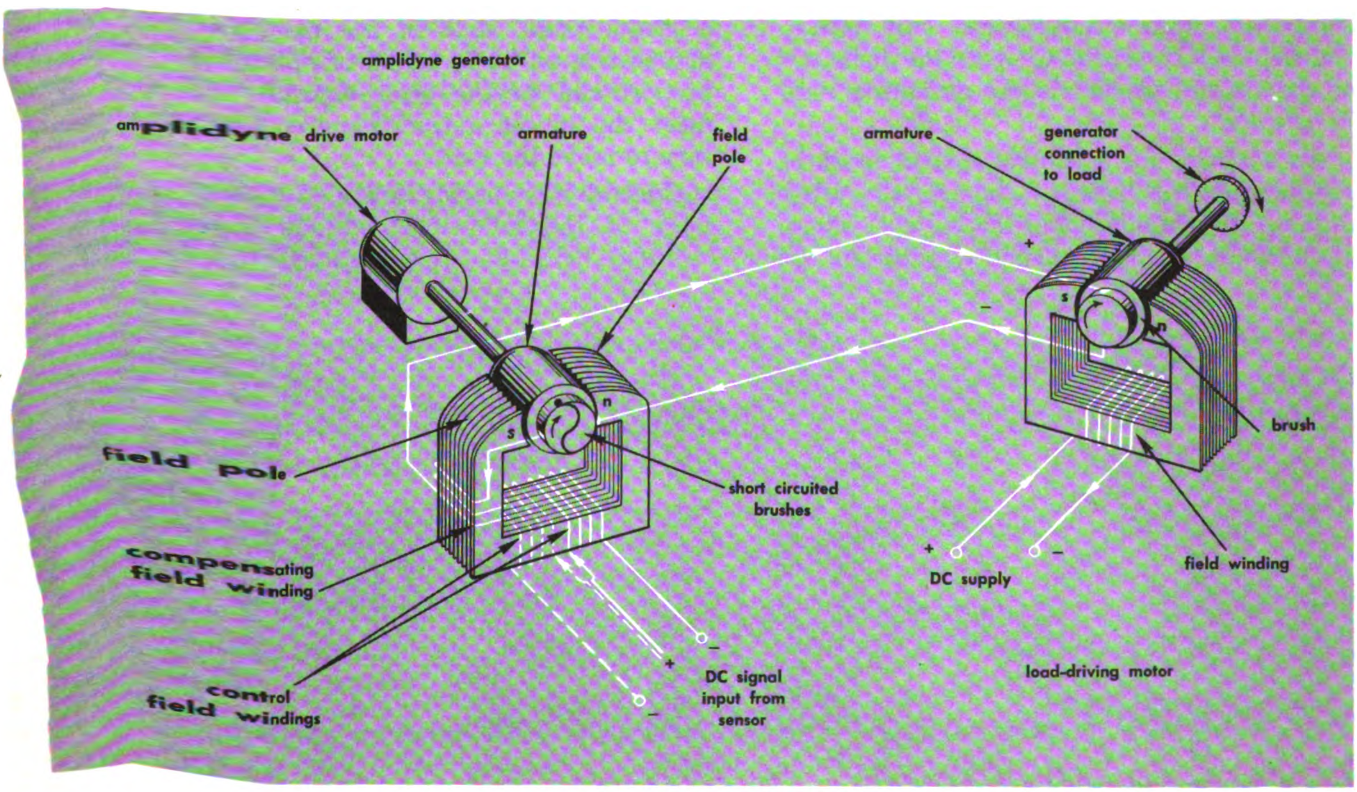
Magnetic fields in an amplidyne

field produced by high currents passing through the shorted circuit. "F_3" is the magnetic field set up by the coils through which the output current flows and which opposes the control field. "F_4" is the magnetic field produced by the output current flowing through the auxiliary windings that have been added to the control field structure. This magnetic field is equal and opposite to "F_3" and therefore balances, or cancels, the effect of "F_3."

The operating principle of an amplidyne has been explained; now we will discuss briefly an application of the amplidyne as represented by the next illustration.

The amplidyne drive motor is connected to a DC supply which causes it to operate at a constant speed and always in the same direction. Therefore, the armature of the amplidyne, which is connected to the shaft of the amplidyne drive motor, is also driven at constant speed and always in one direction. The armature leads of the load-driving motor are connected to the amplidyne generator output brushes. Therefore, the operation of the load-driving motor depends on the DC voltage generated by the amplidyne generator armature. In the illustration, one of the amplidyne control field windings is positioned on the field poles in such a way that if equal currents flow through both windings, the magnetic fields created will cancel and no voltage will be induced in the rotating armature since it is not cutting any lines of flux. However, when a greater current flows through the coil shown on the right (white lines) than through the coil on the left, a magnetic field, having north and south poles as shown, is established. A voltage is generated in the rotating armature. Current then flows through the armature of the load-driving motor causing it to rotate.

After the "controller" block in the missile control block diagram comes the "actuator" block. Actuating units are discussed in the next section.



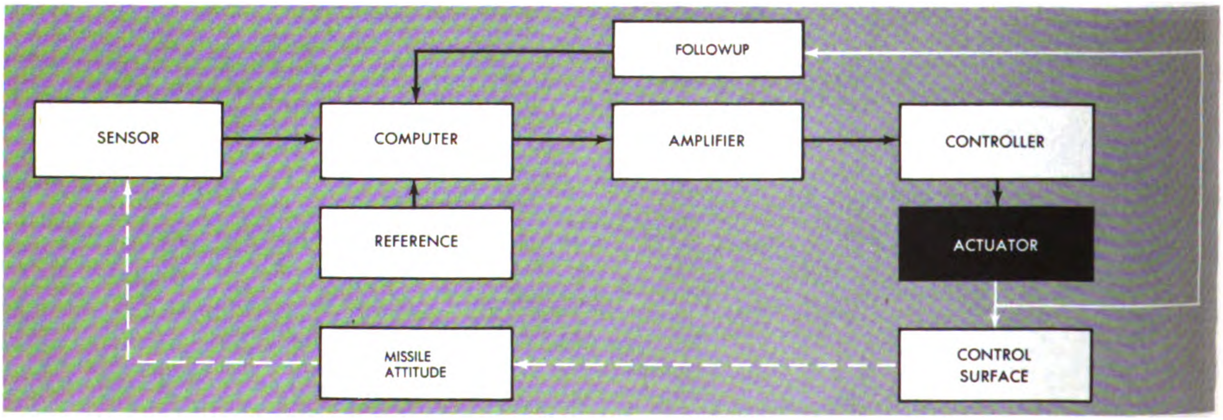
Amplidyne generator connected to a load-driving motor

Actuator Units of Control Systems

In any missile attitude-control system, the energy that originates with detection of an attitude error must be transformed into mechanical motion in order to actuate the appropriate control device. The control system component that accomplishes this energy transfer and transformation at the load end of the control system is the actuating unit, or *actuator*.

An actuator for a control system must be selected on the basis of the characteristics of the other components of the control system. It must be able to respond rapidly to the input error signals. The response involves the time element between reception of the signal and actuation of the control device. At the same time, an actuator must produce an output of proper type and magnitude for a given input; that is, the output of the actuator must be in terms of some function proportional to the error signal and must meet the power output requirement necessary to move the load.

Generally, actuating units employ one or more of the following methods of energy transfer; hydraulic, pneumatic, or electrical.



Basic missile control block diagram

Each method has certain advantages and each also presents certain problems to the designer. The advantages and disadvantages of each method are discussed in chapter 7, which deals with complete control systems.

HYDRAULIC ENERGY TRANSFER UNITS

The underlying principle upon which the transfer of energy by hydraulic means is based is known as Pascal's law. This law states that whenever a pressure is applied to a confined liquid, that pressure is transmitted undiminished in all directions throughout the liquid regardless of the shape of the container forming the confining system.

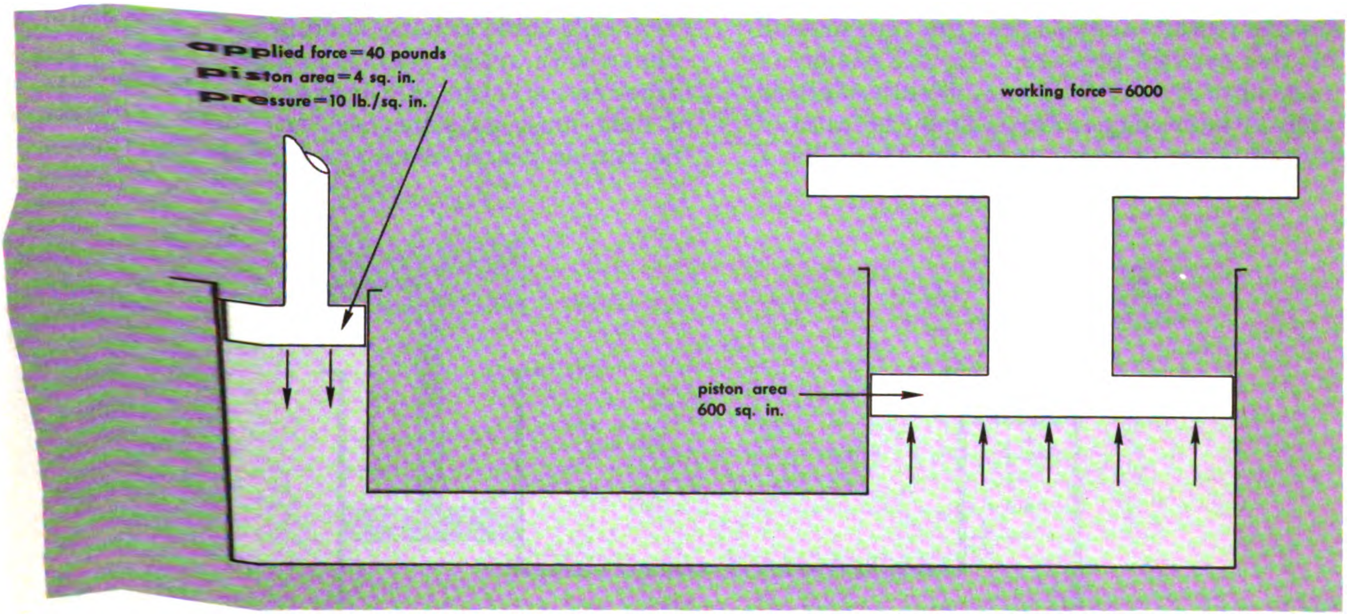
Applications of this principle have been widely and successfully used for many years. A great number of the tools and machines that you have occasion to use from day to day contain some type of hydraulic energy transfer unit. Proof of this statement is found in the hydraulic door stop, hydraulic jack, hydraulic car brakes, hydraulic car lift, and automatic transmission for the present-day auto. Likewise, in the missile industry, applications of this same hydraulic principle are effectively utilized in the transfer of energy through automatic control systems.

Generally, hydraulic energy transfer units are simple in design and construction. The major advantage of a hydraulic system is that it eliminates the use of a complex arrangement of gears, levers, pulleys, etc, for the transfer of energy. Reaction time of hydraulic systems is relatively rapid because there is little slack or play to be taken up as compared to some mechanical systems. The forces generated at one point are transmitted rapidly and with small energy loss over considerable distances. Also, the liquid component (hydraulic fluid) is not subject to breakage, and the complete mechanism is subject to less wear as compared to a totally mechanical system.

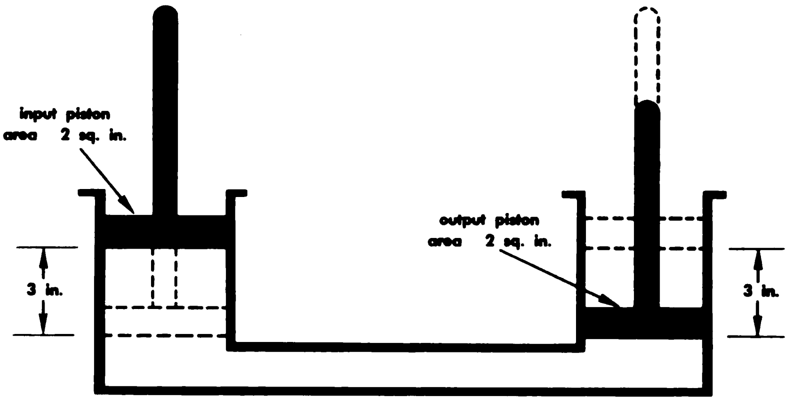
Before discussing any particular hydraulic unit as used in a missile control system, let's consider the construction and operation of a typical hydraulic lift.

Basically, the hydraulic lift consists of a suitable container fitted with two pistons and filled with a fluid which acts as the medium of energy transfer. This construction is shown in the accompanying illustration.

Note that an external force of 40 pounds is applied to the top side of the smaller piston. If the smaller piston has a surface area of



Basic diagram of a hydraulic lift

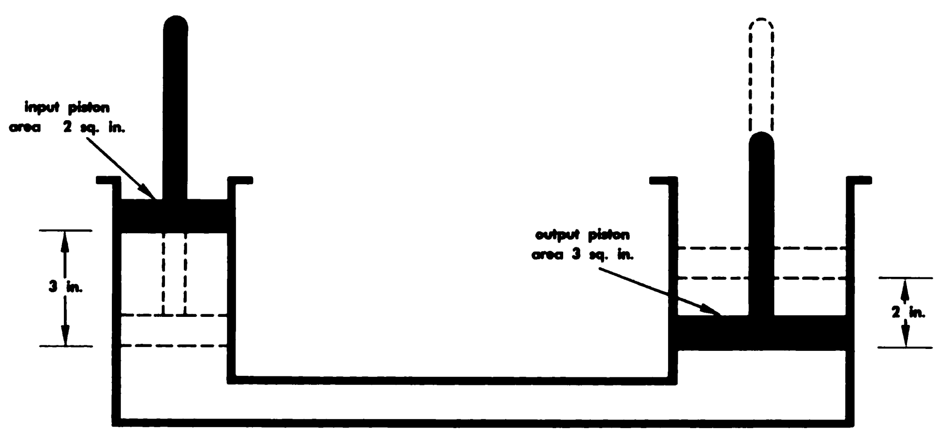


Hydraulic system with equal piston displacement

4 square inches, the 40 lbs. of applied force produces a pressure of ten lbs per square inch on the surface of the confined fluid. The larger piston, upon which the load is placed, has a surface area of 600 square inches. The system pressure of 10 pounds per square inch acts on each square inch of the large piston, thus producing a working force of 6000 pounds. We may conclude, then, that in addition to being used for the transfer of energy, hydraulic systems can also be devised to produce a large output force with the expenditure of a much smaller input force.

It is also important to note that the varying size of the container housing the fluid and the irregular path of the delivery line do not (within limits) affect the transmission of the applied pressure. Another factor to remember is that the ratio of output piston area to input piston area equals the ratio of output force to input force. In short, for a given pressure, the force produced by the working piston is directly proportional to its area.

In certain applications of hydraulic systems, the distance moved by the output piston may be of primary importance instead of force amplification. For example, it may be desirable to have the output piston move some predetermined distance. This factor can be controlled by using pistons with properly related surface areas. Consider the hydraulic system above. Both pistons have surface areas of 2 square inches. If the force applied to the input piston causes it to move a distance of 3 inches, then 6 cubic inches of fluid have been displaced. In this instance, 6 cubic inches of fluid is a cylindrical column with an end surface area of 2 square inches and a height of 3 inches. This column of fluid must go someplace, so it displaces the output piston. Since the surface area of



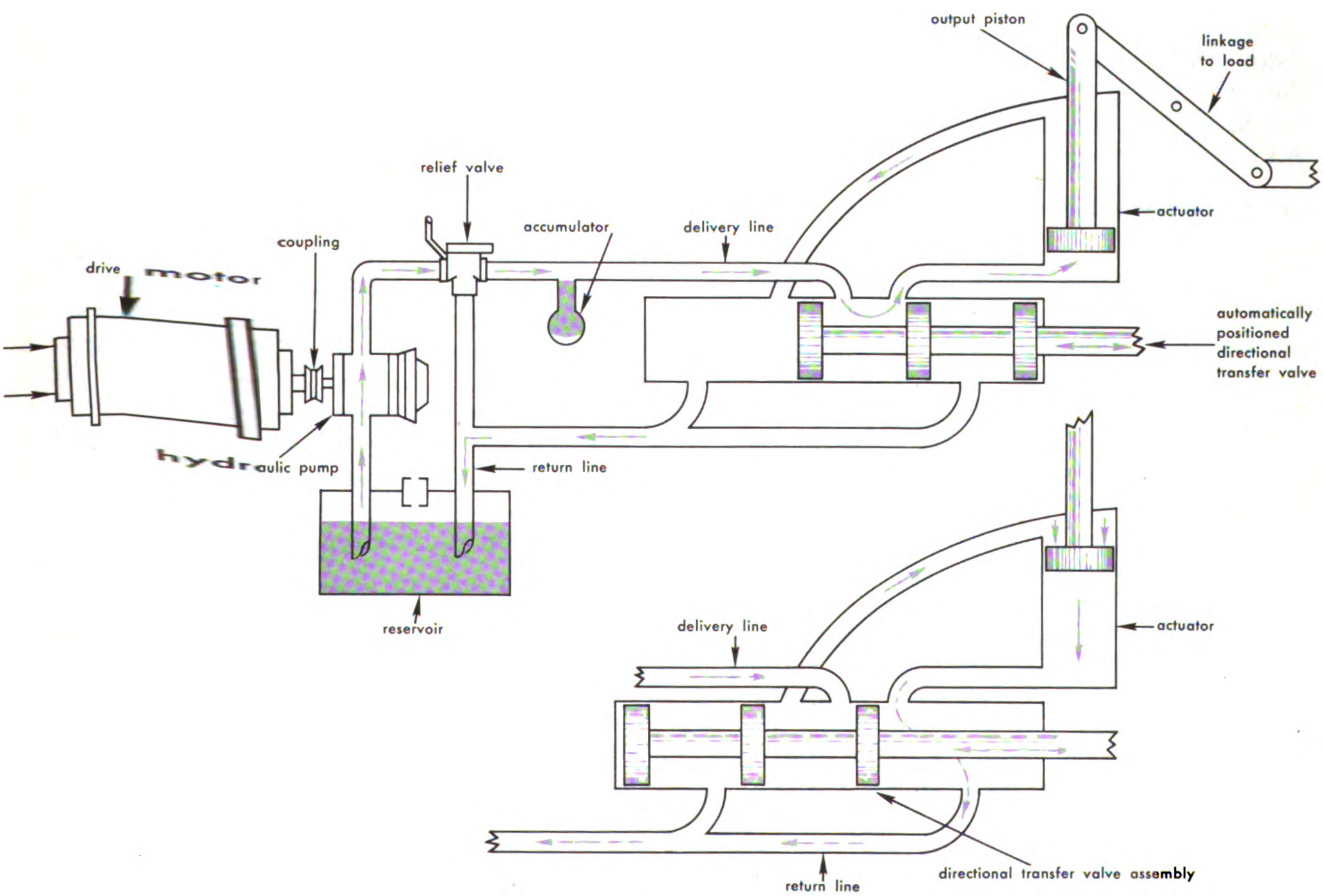
Hydraulic system with proportional piston displacement

the output is 2 square inches, it must be displaced a lateral distance of 3 inches in order to make room for the six cubic inches of fluid.

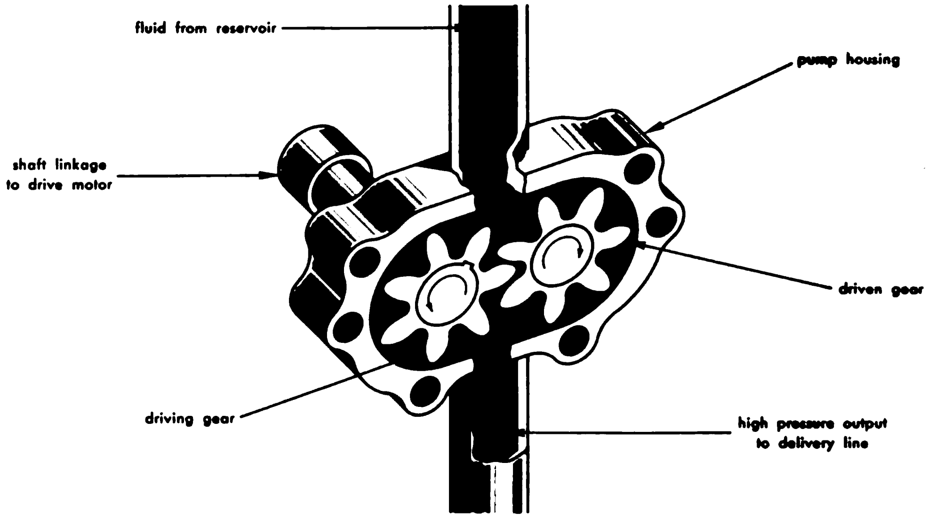
Now consider the preceding diagram in which the surface area of the input piston is 2 square inches and the surface area of the output piston is 3 square inches. If the force applied to the input piston causes the piston to be displaced a distance of 3 inches, 6 cubic inches of fluid will be forced into the output cylinder. The output piston, having a surface area of 3 square inches, will move only 2 inches to make room for the 6 cubic inches of fluid. Thus, neglecting frictional losses within a hydraulic energy transfer system, the input force *times* the distance through which it moves *equals* the output force *times* the distance through which it moves *equals* the output force *times* the distance through which the force acts. This force-distance relationship is an important factor in the design of hydraulic actuating units.

Now consider the application of pressure transmission in fluids to the actuation of attitude control devices in missiles. Depending on the work to be performed, the parts of a hydraulic actuating system will vary in arrangement and physical dimensions. However, the illustration of a simple hydraulic system below shows the relative location of the basic components. The direction of fluid flow through the system is indicated by the arrows.

In reference to the illustration, power supplied by the electric motor drives the hydraulic pump. The pump unit forces fluid under pressure through the delivery line to the directional control valve. The automatically positioned directional transfer valve



Components of a simple hydraulic system



Gear-type hydraulic pump

determines the direction of fluid flow into the piston-type actuating unit. By means of the actuating unit arrangement, the motion of the piston and the force acting on the piston are transmitted to the load (control device) by the mechanical linkage. Let's discuss individually the basic components of this hydraulic system.

Construction and Operation of Hydraulic Pumps

The primary energizing unit of the hydraulic system is the force pump. This pump can be driven by an electric motor or by some other energy source within the missile. This energy source must supply energy proportional to the attitude error. Two types of frequently used pumps are the geared-type and the piston-type (reciprocating pump).

The geared-type pump consists mainly of two tightly meshed gears which revolve in a housing as illustrated above. The clearance between the teeth of the gears and the housing is small. For operation, the intake port is connected to the fluid reservoir, and the output port is connected to the high-pressure delivery line of the hydraulic system. As the gear teeth pass the edge of the intake port, fluid is trapped between the teeth and the housing. This fluid is carried around the housing to the output port. As the gear teeth mesh in front of the output port, the fluid between the teeth is forced into the high-pressure delivery line. The hydraulic actuator and the load to which it is attached are displaced by the fluid being pumped into the system.

As mentioned, the second type of force pump that can be used in a hydraulic energy transfer system is the reciprocating or piston-type pump. Pumps of this type depend on the back-and-forth motion of moving parts, working in conjunction with suitable valves, to force the hydraulic fluid into the high-pressure line. One advantage of the piston-type pump is its ability to develop higher pressures (relatively speaking) than other types of pumps. The illustration at the right shows a simple double-acting pump which produces a continuous flow of fluid. A greater quantity of flow and fewer pulsations of flow can be achieved somewhat by increasing the number of pistons in a piston-type pump.

Note in the diagram that the piston can move in both directions (double-acting). As the piston moves from right to left, a partial vacuum tends to develop in the right-hand section of the chamber. Immediately, fluid under atmospheric pressure in the reservoir forces valve 1 open and enters the chamber. At the same time, fluid on the discharge side of the pump attempting to reenter the chamber closes valve 2. Thus, the space left vacant by the piston as it moves from right to left is being charged with fluid from the

reservoir. Also, as the piston moves from right to left, it exerts a force on the fluid in the left-hand section of the chamber, thus closing valve 3 and opening valve 4. Fluid passes through valve 4 to the pressure line. When the piston reaches the extreme left position, the chamber section to the right of the piston is completely filled with fluid. On the return stroke, from left to right, valve 4 closes and valve 3 is forced open by fluid from the reservoir which enters the chamber to fill the space left vacant by the piston. At the same time, valve 1 closes, valve 2 opens, and fluid is forced into the delivery line.

To summarize this action, on the right-to-left piston stroke, fluid enters valve 1 and is discharged through valve 4, while valves 2 and 3 are closed. On the left-to-right piston stroke, fluid enters valve 3 and is discharged through valve 2, while valves 1 and 4 are closed.

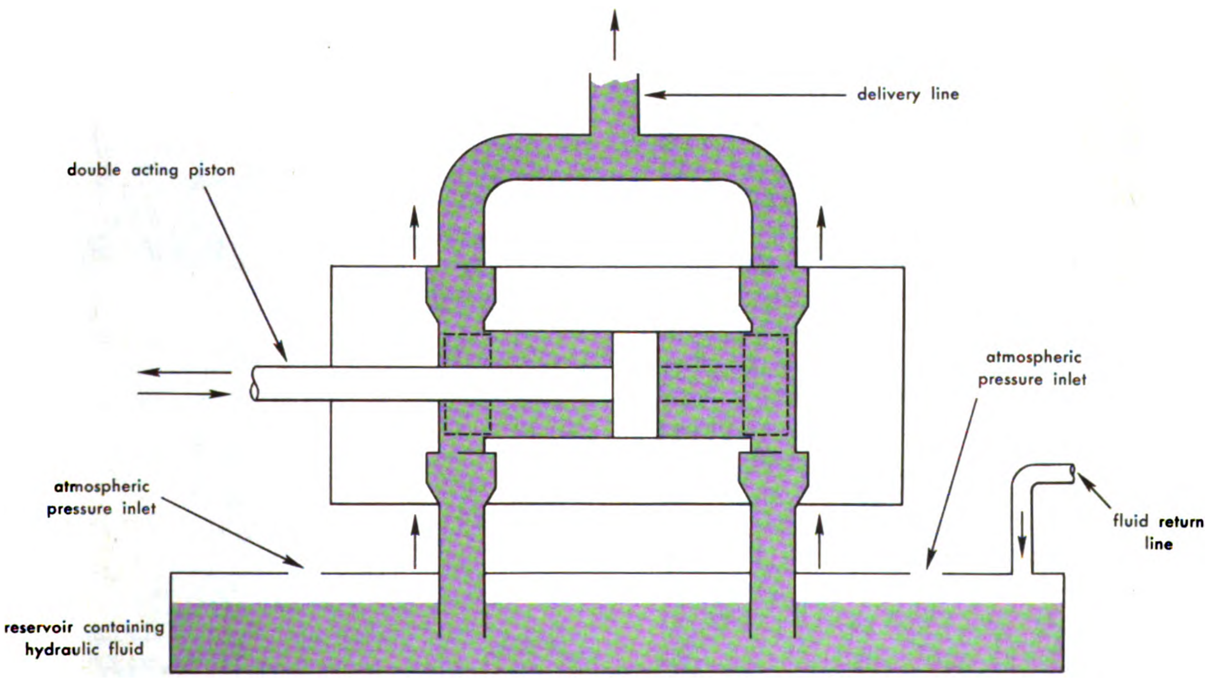
Function of a Reservoir in a Hydraulic System

A reservoir is the storage place for the fluid used in a hydraulic system. Fluid flows from the reservoir to the pressurizing pump which forces the fluid throughout the hydraulic system. The reservoir also receives the returning fluid after it has performed the desired work on the hydraulic actuator piston.

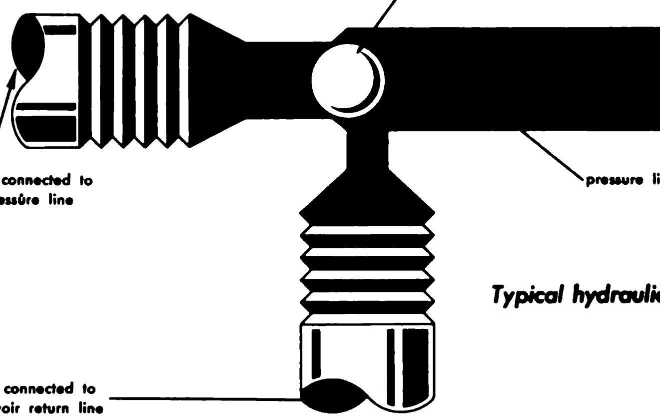
Construction and Operation of a Typical Relief Valve

Pressure-relief valves are designed to limit the pressure in a hydraulic system to some maximum value and thus prevent damage to parts of the system. Some hydraulic systems use hydraulic pressure-regulating switches instead of relief valves. Such switches control the power-pump operation in response to pressure changes within the system.

The typical pressure-relief valve shown on the next page consists of a metallic housing with two ports. One port is connected to the pressure line of the hydraulic system. The other port is connected to the reservoir return line. Notice that the spherical ball is held seated in the restricted section of the pressure-line by a spring, which prevents fluid from passing to the reservoir return port under normal operating conditions. However, when the pressure at the pressure-line port becomes great enough to overcome the



Double-acting piston-type hydraulic pump



Typical hydraulic pressure relief valve

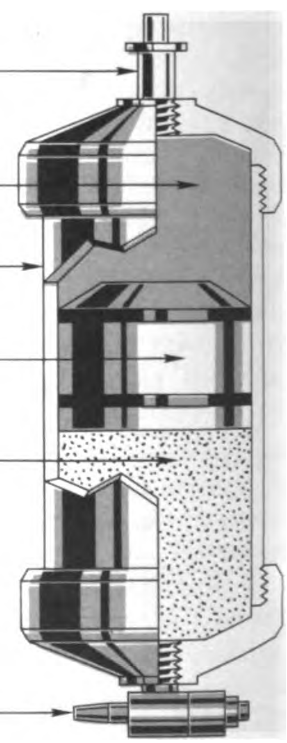
force exerted on the ball by the spring, the ball is moved off its seat. This action allows fluid to escape through the reservoir return port and reenter the reservoir. Thus, the pressure in the hydraulic system can never go above the value necessary to overcome the force exerted by the spring designed for that particular system.

Construction and Operation of a Hydraulic Accumulator

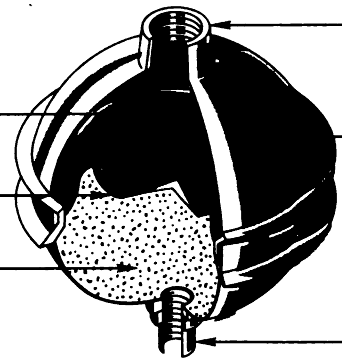
An accumulator acts as an auxiliary storage place for hydraulic fluid under pressure. In so doing, it tends to dampen out pulsations or pressure surges in the hydraulic system. Pulsating flow in a hydraulic system would cause vibration of components and unsteady operation of the control devices to which the actuators are linked.

Accumulators may be of either the floating piston type or the diaphragm type, both of which are shown in the following illustrations.

The floating piston-type accumulator consists of a metallic cylinder separated into a hydraulic fluid chamber and an air chamber by the floating piston. The diaphragm type consists of two hollow hemispherical pieces of metal separated into a hydraulic fluid chamber and an air chamber by a flexible diaphragm. In both types, the air chamber is charged with compressed air to a pressure corresponding to the line pressure desired in the hydraulic system which exerts a force on the piston, or the diaphragm. If the line pressure builds up higher than the air pressure in the accumulator, fluid is forced into the hydraulic fluid compartment. This fluid pushes the piston down, or depresses the diaphragm, thus further compressing the air in the air chamber. During periods of peak load, or power-pump lag, the compressed air tends to force fluid from the accumulator back into the hydraulic system. Thus, by building up pressure in an accumulator, variations in hydraulic-system line pressure are smoothed out.



Floating piston-type hydraulic accumulator



Diaphragm-type hydraulic accumulator

Function and Operation of a Hydraulic Actuator

The purpose of a hydraulic actuator is to transform fluid pressure into mechanical force necessary for moving some type of control device. A basic hydraulic actuator consists essentially of a cylinder, with suitable fluid intake and exhaust ports, fitted with a piston and connecting rod. The actuator shown in the illustration of a simple hydraulic system on page 249 is a double-acting piston type. In the actuator, the pressurized hydraulic fluid can be applied to either side of the actuator piston, thus producing motion in either of two directions. On the upstroke, fluid under pressure enters the cylinder below the piston, forcing the piston up and forcing the fluid above the piston back to the reservoir. The downstroke of the actuator piston results when fluid enters the cylinder above the piston. In this instance, the fluid below the piston returns to the reservoir.

The second method of energy transfer that we'll discuss involves the use of pneumatic units.

PNEUMATIC ENERGY TRANSFER SYSTEMS

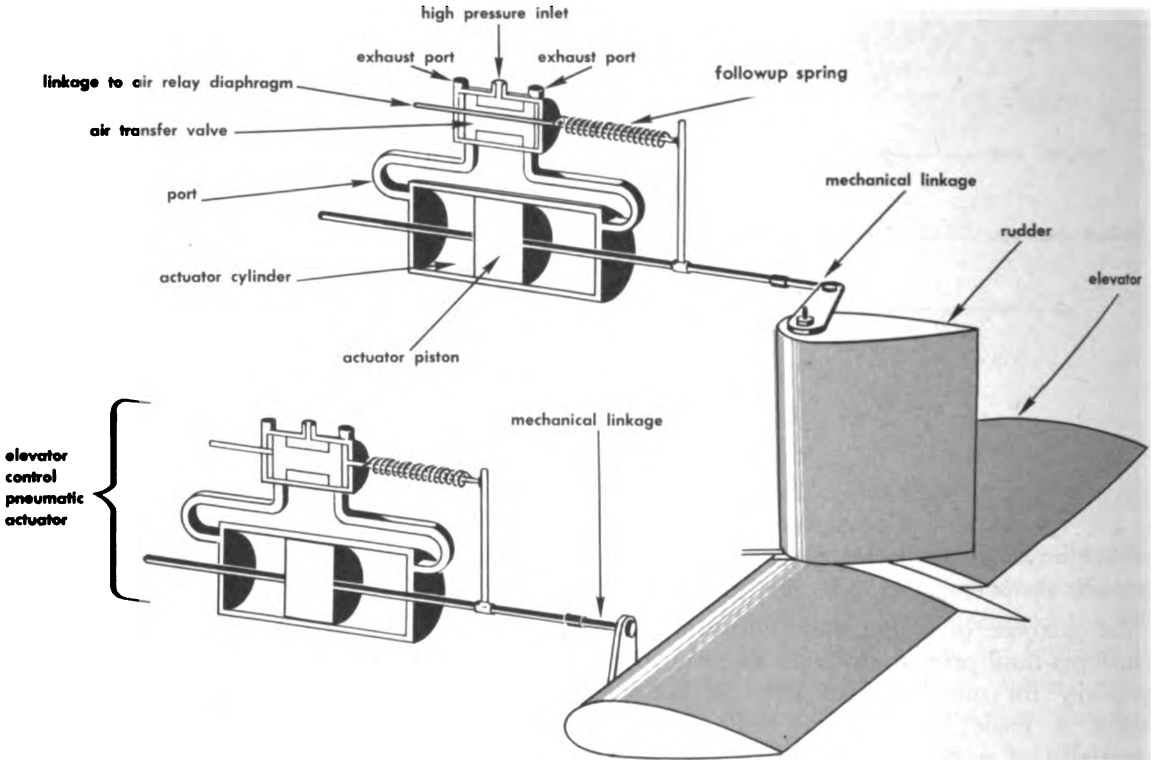
Energy transfer systems that use air as the medium of energy transfer are referred to as pneumatic systems. Basically, the operation of a pneumatic system is similar to the hydraulic system just discussed. The most prominent difference between the two systems is that the medium of transfer in the pneumatic system is a gas while the medium of transfer in the hydraulic system is a liquid.

In a pneumatic system, air from a pressurized source passes through suitable delivery tubes, valves, and pressure regulators to do work upon some mechanical unit. The pressurized source generally consists of high-pressure air stored in metal tanks. The pressure energy, originally possessed by the air, is transferred from one point to another where it is transformed into mechanical work by a piston, or diaphragm, which is connected to the missile control device.

Unlike a hydraulic system, a pneumatic system does not re-use its transfer medium after it has performed work on the load. For that reason, the air must be stored at a pressure much higher than that necessary for actuating the load in order to maintain adequate system pressure as the stored air supply diminishes.

The accompanying diagram on the next page shows two double-acting piston-type pneumatic actuators used for rudder and elevator control.

Refer frequently to this diagram as you read the following description of its construction and operation. The air transfer valve is mechanically linked to an air relay (not shown) which receives the air error signal produced by deviations in azimuth of the missile. The relay action is such that both force and sense of the error signal are transmitted to the air valve. Thus, the initial direction of displacement of the air transfer



Pneumatic actuating units mechanically linked to control surfaces

valve is determined by the sense of the air error signal. Assume that an azimuth deviation has occurred, and that the error signal from the air relay has caused the air valve to move toward the right, thus opening the right port to the actuator cylinder. Air from the high-pressure inlet passes through this port and causes the actuator piston rod to move toward the left. Simultaneously, air is pushed from the left-hand section of the actuator cylinder through the port located in that section. This displaced air is exhausted into the atmosphere through the air-transfer-valve exhaust port located at the left end of the air transfer valve.

The actuator piston motion is transmitted through the mechanical linkage to the rudder, which applies corrective control in the proper direction to bring the missile back to its neutral position in azimuth. As the piston moves, it exerts a force on the followup spring. The followup spring is a calibrated coil spring connected between the piston rod and the air transfer valve. Movement of the piston puts the spring in a state of tension or compression, depending on the direction of piston displacement. In either state, the followup spring exerts a force on the air transfer valve which opposes the force exerted by the air relay. Thus the air-transfer-valve movement is the difference of two forces and is in the direction of the resultant force. Movement of the actuator piston continues until the force exerted by the followup spring is equal, but opposite in direction, to the force exerted on the air transfer valve by the air relay. When this condition is established, the air-transfer-valve spool is centered, and actuator piston movement stops. With the air transfer valve thus balanced, air leaks past the air-transfer-valve spool and puts equal pressure on both sides of the actuator piston. This action holds the piston and the rudder to which it is linked in the corrective positions ordered by the air error signal.

The force applied to the rudder causes the missile to return toward straight flight attitude. Since the missile is now moving in a

direction which is opposite to that during off course displacement, an opposing air rate signal is applied to the air relay. This air signal reduces the force that the air relay is exerting against the air transfer valve. As the missile approaches normal attitude in azimuth, the force exerted by the followup spring is greater than that exerted by the air relay. Now the air valve moves to the left, and air from the high-pressure inlet forces the actuator piston toward its neutral position; that is, the direction of motion of the piston is now from left to right. This action results in partial removal of the control force applied to the rudder. As the amount of control is reduced, the force exerted by the followup spring is reduced. Therefore, the rate of turn and consequently the rate signal are also reduced. Thus, all forces are continually being reduced as the missile approaches normal attitude. When normal attitude is regained, the air signal is zero, the followup spring force is zero, and the actuator piston and rudder are again centered. All movement ceases until the missile again deviates in azimuth due to air gusts or its own flight characteristics.

This same type actuating unit could be used to position other control devices with proper modification of the mechanical linkage. For example, the elevator control actuator, shown in the preceding diagram, is exactly the same except for the manner in which the mechanical linkage is connected to the elevator.

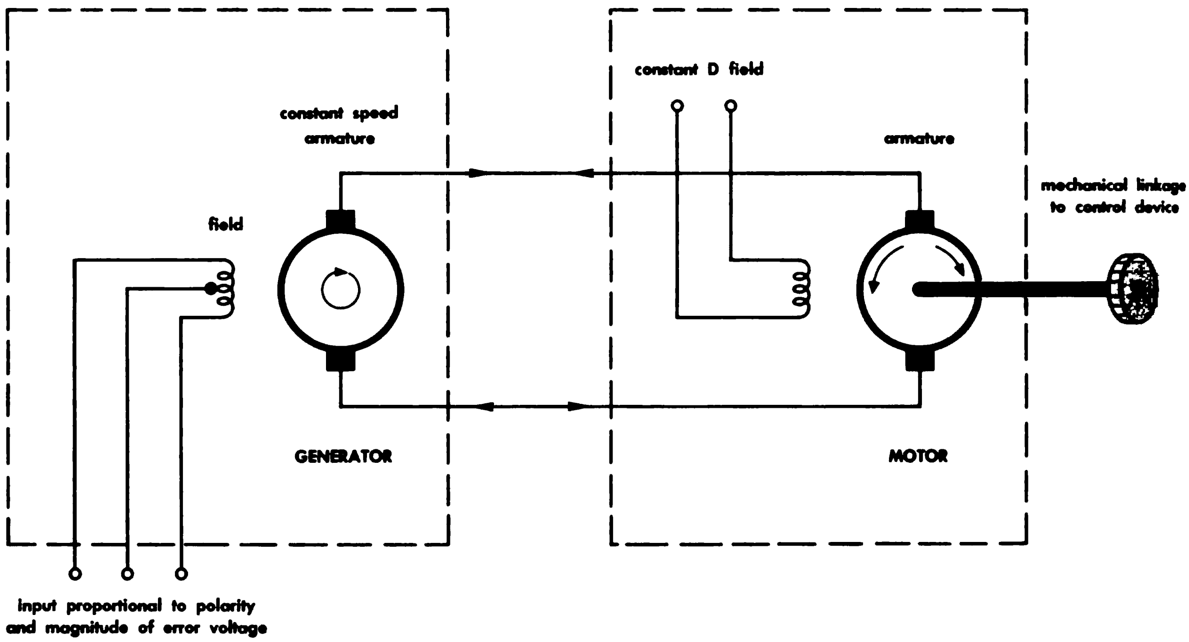
ELECTRICAL ENERGY TRANSFER SYSTEMS

Electrical energy transfer systems are numerous. The assembly of electrical components into systems for transferring energy varies widely, depending on the job the system must perform in a missile. However, in any case, the objective of such a system relative to actuating control devices is to transfer electrical energy from one point to another where it is transformed into mechanical motion of a control device. Generally, motors are used as the actuators in electrical energy transfer systems.

The type of electric motor used as an actuator depends primarily on the size of the load and the speed with which the load must be moved. In general, DC motors develop higher stall torque than AC motors and, therefore, are more often used for driving heavy loads of the type encountered in high-speed missile control. An AC motor is inherently a constant-speed device and thus not suitable for the requirements of a servo motor where variation in speed of rotation is necessary. This factor also makes the DC motor more applicable than the AC motor as an electric actuator.

Since the armature of the DC motor is mechanically linked to the load, some means of controlling the speed and direction of rotation must be utilized. Two ways of accomplishing this control are by controlling the voltage applied to the field coils and controlling the armature voltage. Controlling the field voltage would permit control of armature speed only, upward from a certain minimum speed. In order to permit the speed of rotation of the armature to be zero at times, the armature method of control must be used. Thus, if no current flows through the armature windings, no rotation is produced. On the other hand, when current does flow through the armature windings, the speed and direction of rotation will be determined by the magnitude and phase of the error voltage. Usually, the servo motor armature voltage is controlled by feeding the error voltage into the field of a DC generator, the output of which is fed to the armature of the DC motor.

The diagram on page 256 shows the schematic arrangement of the components of a DC servo motor. The DC generator armature is driven at a constant speed of rotation. When the DC generator field is energized by the amplified error signal current, a voltage is induced in the armature coils which are cutting the magnetic field. The output voltage induced in the armature coils is therefore proportional to the speed of rotation of the generator armature and the magnetic field strength. Since the armature is driven at constant speed, the output voltage of the generator varies with the magnetic field. The magnetic field is, in turn, proportional to the magnitude and polarity of the signal voltage used to excite the field. The variable output of the generator is fed to the armature of the DC servo motor. The field of the DC servo motor is excited by a constant DC



Basic schematic of a DC motor used as an actuator

source. Thus, the direction and speed of rotation of the armature is proportional to the polarity and magnitude of the armature current. The armature of the DC motor is mechanically linked to, and drives, the control device.

Utilization of a DC motor as an actuation unit controlled by a generator as shown in the diagram on the opposite page produces an undesirable reaction which must be eliminated if servo control is to be efficient. The problem arises because of the tendency of the magnetic structure of the generator to remain magnetized. It tends to remain magnetized after the magnetizing current has been diminished to zero by correction of the control misalignment that produced that current. The residual magnetic field tends to induce a small output voltage in the rotating armature coils even during the absence of field excitation current. This reaction is extremely undesirable for servo applications, because the DC servo motor will continue to develop torque (rotate) in the absence of an error signal. A reaction of the type just described causes the load to move farther than desired.

A solution to this problem can be achieved by using additional windings on the generator field structure. If these auxiliary windings are excited by low-level AC, the alternating magnetic field which results causes the average magnetization of the generator field to fall to zero when there is no excitation current flowing in the main field winding. However, you can see that the AC flowing in the auxiliary winding also causes small fluctuations in the main magnetic field. This reaction, in turn, produces a small AC component in the generator output to the servo motor.

As a solution to this secondary problem, instead of the additional windings for inserting an alternating magnetic field, small permanent magnets are mounted on the generator armature in such a way as to revolve with the armature. This arrangement of permanent magnets provides an alternating magnetizing force for the generator field and permits the DC output of the generator to fall to zero when the main field is not excited. Since the permanent magnets have a fixed position relative to the armature coils, they cannot induce voltage in these coils. Therefore, the generator output is free from an AC component and the servo motor torque is more accurately controlled by the error voltage.

COMBINATION-TYPE ENERGY TRANSFER SYSTEMS

In some instances, actuators for control devices employ two different methods of energy transfer to achieve the desired result. For example, an electro-pneumatic actuator consists essentially of a pneumatically operated piston whose direction of motion is determined by an electrically positioned selector. Actually, this device could be considered a controller-actuator combination; that is, the electrically positioned valve acts as the controller, and the pneumatically driven piston is the actuator. However, the compact arrangement of the components of this combination-type actuator justifies placing a discussion of it in this section.

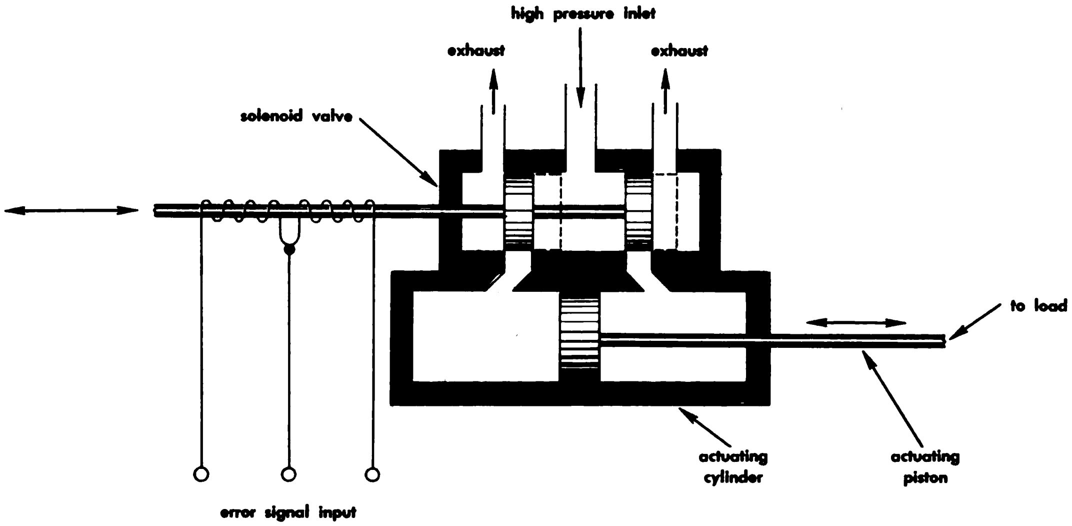
The accompanying diagram illustrates the arrangement of the components of a typical electro-pneumatic actuator unit.

Notice that the solenoid valve controls the intake and exhaust of air to and from the actuating cylinder. The position of this valve is determined by the polarity of the error signal input. Also, the length of time that the valve remains in a certain position depends on the time interval during which the error signal of a certain polarity exists.

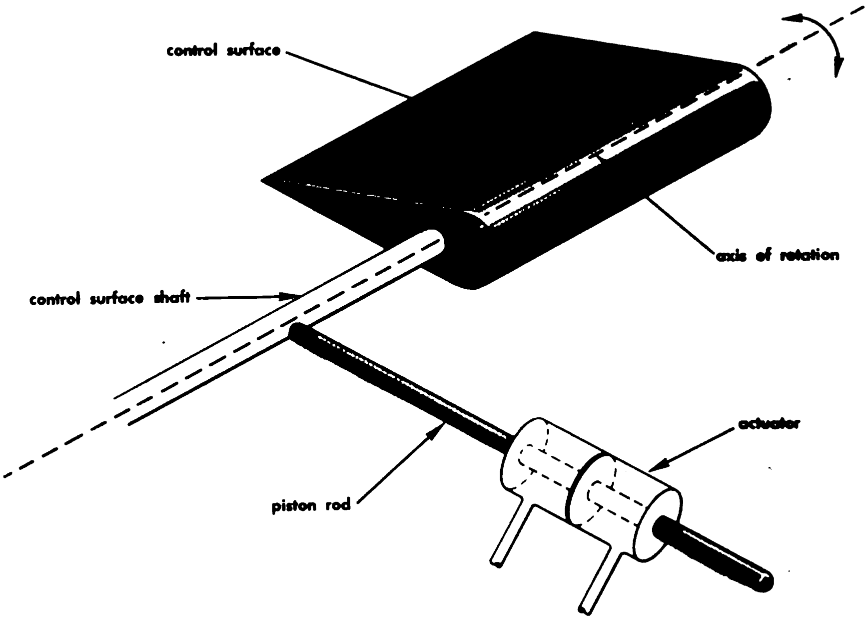
As illustrated below, the valve and piston occupy neutral positions. Now, assume that an error signal of a given polarity exists that will cause the valve to move toward the right, but not far enough to exceed the displacement limits indicated by the dotted lines. This new valve position permits high-pressure air to pass through the valve into the actuating piston. The actuating piston is forced toward the left and, at the same time, air on the left side of the piston is forced out through the valve by way of the left exhaust port. During this action, the exhaust port located on the right is closed. When the attitude error has been corrected, the valve and actuating piston will again be in netural positions.

When an error signal of opposite polarity exists, the solenoid valve moves toward the left. High-pressure air now enters the actuating cylinder on the left side of the piston and forces it toward the right. Air forced from the right section of the actuating cylinder passes through the valve and escapes by way of the right exhaust port.

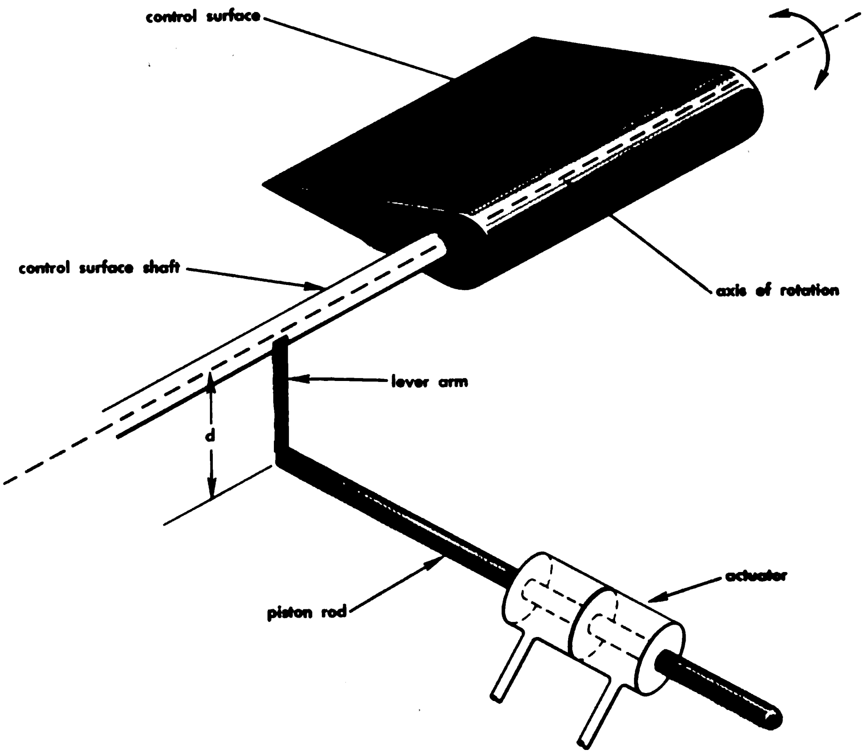
The basic construction and operation embodied in an electro-pneumatic actuating unit can also be employed in an electro-hydraulic actuator. The major change would be the use of hydraulic fluid instead of air for driving the load actuating piston. In either case, the solenoid valve permits rapid response of the actuating unit when an error signal exists.



Basic diagram of an electro-pneumatic actuator unit



Actuator linked to load without lever arm



Actuator linked to load through a lever arm

USES OF MECHANICAL LINKAGES FOR ENERGY TRANSFER

In order to transfer energy effectively from the actuator to the control device in an attitude control system, it is necessary to use some type of mechanical linkage. With proper designing, a mechanical linkage (in addition to transferring the energy) also permits a more advantageous application of a given force to the control device. This advantage is accomplished by changing the point of application as well as by producing an amplification of the torque moment.

To further explain these statements, let us consider a situation in which a hydraulically operated piston is being used to actuate a control surface designed to rotate about an axis as illustrated on the left.

With the piston rod connected to the diametric center of the control-surface shaft as shown, no rotation of the shaft results, regardless of the force exerted by the actuator.

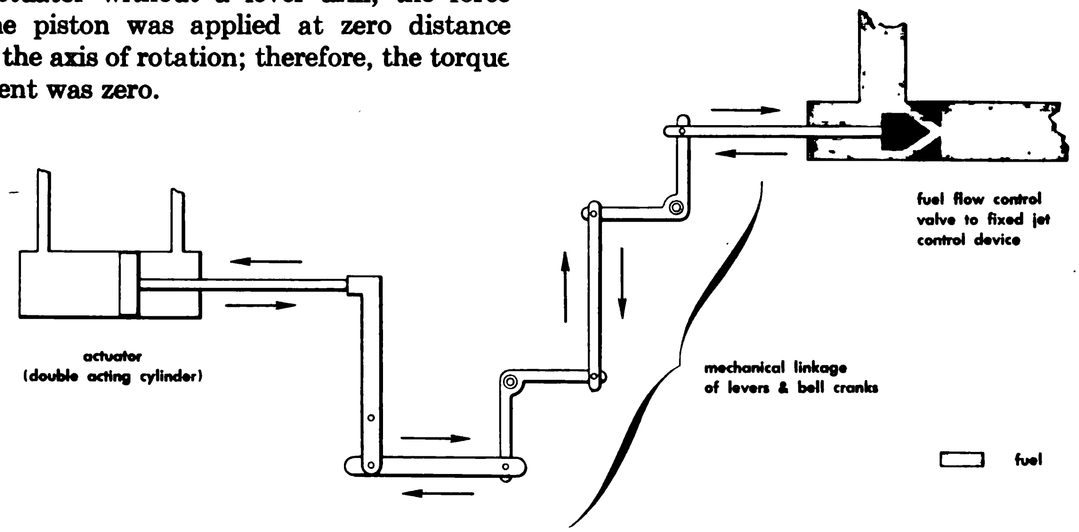
But, suppose we apply the actuator output force to a lever arm fixed to the control-surface shaft and let the point of application be some certain distance (d) from the axis of rotation. This type of arrangement is shown in the figure on the left. Now the control surface rotates because the force exerted by the piston is applied at a distance from the axis of rotation and produces a torque moment for a given force. In the diagram of an actuator without a lever arm, the force of the piston was applied at zero distance from the axis of rotation; therefore, the torque moment was zero.

Mechanical linking devices generally consist of an arrangement of gears, levers, and/or cables. Of course, the linkage must be arranged in such a manner that the sense and magnitude of the actuator output will cause movement of the control device, of proper magnitude and direction, to correct the existing attitude error. Reversals in actuator output produce reversals in the direction of motion of the mechanical linkage components. Such reversal motion, in turn, causes a reversal in the direction of motion of the load.

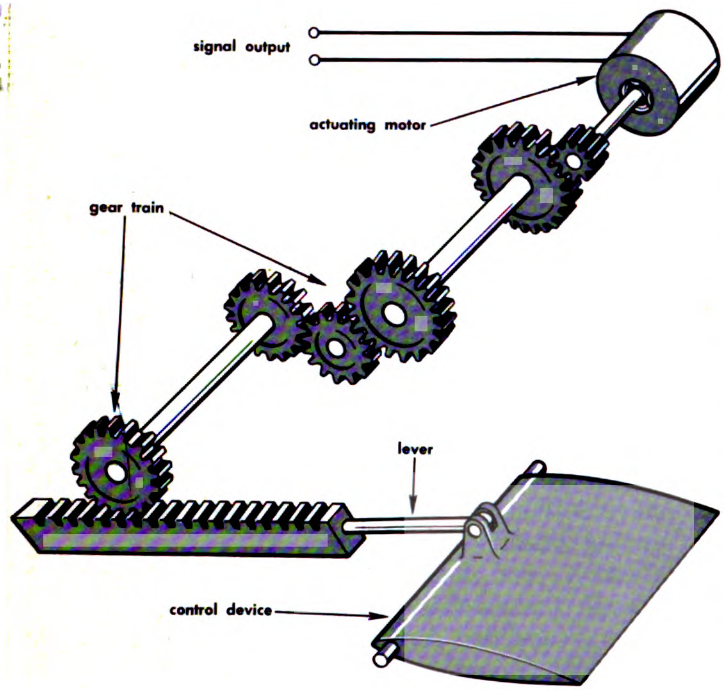
As a summing up of this section on mechanical linkages for energy transfer, study the following illustrations. They show various possible arrangements for mechanically linking the actuator to the load. In each case, imagine that the actuator is moving (either moving linearly or rotating) in some given direction, and follow the resulting motion through the mechanical linkage to the load. Then reverse the direction of the motion of the actuator and again mentally follow the action of the mechanical linkage and direction of displacement of the load.

RUNDOWN OF ENERGY TRANSFER METHODS

As you know, the method of energy transfer used in an actuating unit depends on the other components of the control system. The merits of each method are considered in Chapter 7.



Levers and bell cranks used as a mechanical linkage to actuator and load

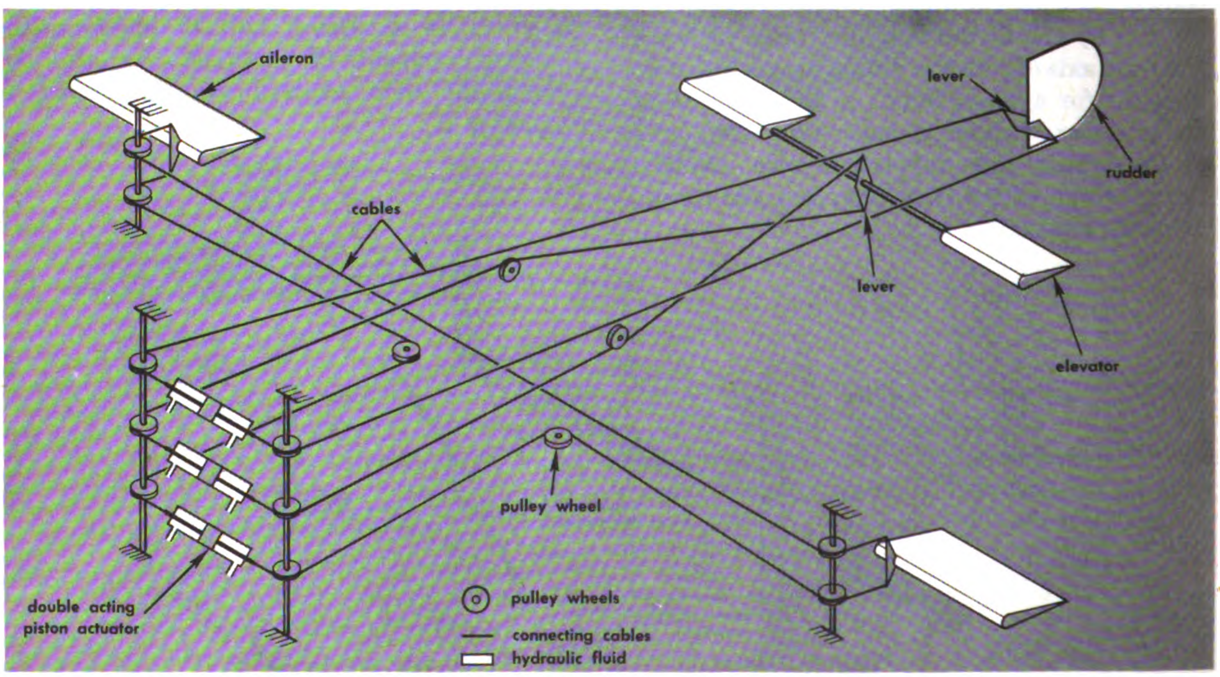


Gear train type of mechanical linkage

The transfer of energy by hydraulic units is based on Pascal's law which states that whenever a pressure is applied to a confined liquid, that pressure is transmitted undiminished in all directions throughout the liquid regardless of the shape of the container forming the confining system. Basically, a pneumatic system operates in a manner similar to a hydraulic system. A principal difference of the two systems is the medium used for transfer of energy. Liquid is used in hydraulic systems, while air is the medium in pneumatic systems.

The assembly of electrical components in missile electrical energy transfer systems varies widely, according to the job the system must perform. These systems generally employ DC motors as actuators. Sometimes, two methods of energy transfer are employed in an actuator unit. An electro-pneumatic actuator is such a combination unit. It consists of a pneumatically operated piston whose direction of motion is determined by an electrically positioned selector.

Energy transfer by mechanical linkages in missiles is generally done by an arrangement of gears, levers, and/or cables.



Cable linkage from actuators to control surfaces

Followup Units of Control Systems

A followup unit of a missile control system provides continuous information on control-surface position in relation to the airframe by means of a *followup signal*. This information is usually called *followup* since it indicates how the output of the control system is *following* on error signal.

FOLLOWUP SIGNALS INDICATE OUTPUT OF CONTROLLING EQUIPMENT

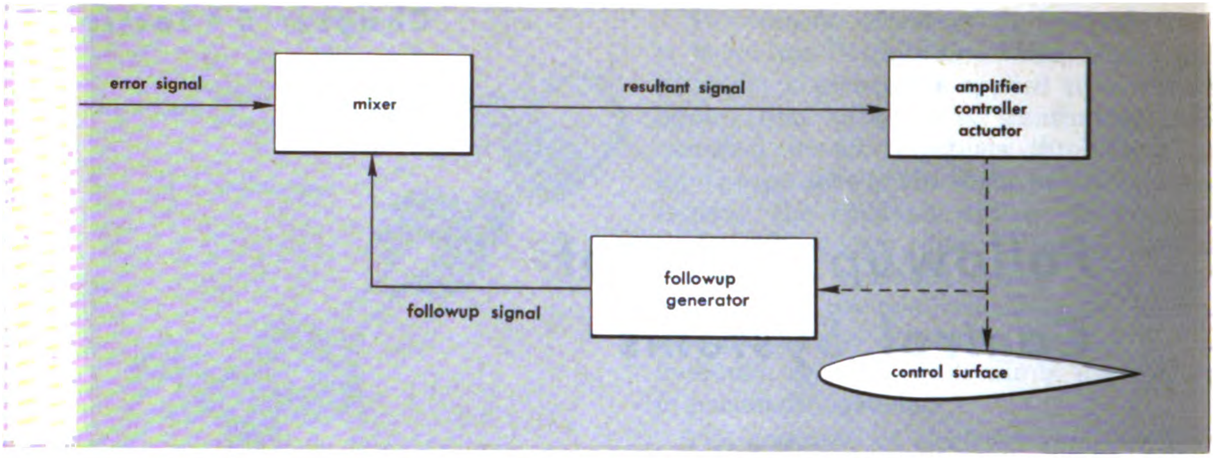
A followup signal is an indication of the output of the controlling equipment. The signal is approximately proportional to the surface deflection from streamline. It is an indication of the output of the controlling equipment. Missile deviation is represented when the signal combines with the error signal.

Without a followup signal there would be nothing but the varying air pressure to prevent the control surface from swinging its limit and "hitting the stops" any time an error signal from the sensor existed. The followup enables a surface to deflect an amount which is dependent on error signal strength. It also provides a signal to return the surface to streamline.

The followup signal combines with the error signal in a manner that *opposes* the error signal. The error signal is large enough, however, to produce the necessary surface deflection. When the followup becomes as large as the error signal, the surface does not deflect any farther since the resultant of the two signals is zero. (The resultant is the signal that operates the control surface.) If an error signal is strong, the surface deflection is large before the signals exactly counteract. Since the missile then starts returning to the desired heading, the error signal becomes less. The followup signal is then the larger, so the resulting voltage is reversed. This reversal moves the control surface in the *opposite* direction, until it is again streamlined. Such action provides smooth and rapid control which cannot be imitated by an "on-off" system which affords intermittent control and overcontrolling.

FOLLOWUP LOOPS SUPPLEMENT MAJOR FEEDBACK SIGNALS

A followup loop supplements the major feedback signal of a control system. The followup loop has a generator which creates a feedback signal that is in addition to the major dynamic feedback path. The *major* path represents information on the reaction of the missile attitude fed back to the sensor. The angular movement of the aircraft completes this major loop. The *minor* feedback path (followup loop) returns information on the reaction of a *control surface* rather than the reaction of a missile. This return completes a minor loop which includes less equipment and is of less importance, in principle, than the major loop. Both loops are included in the basic missile control block diagram shown on page 246. The followup loop is shown in the following figure.



Followup loop of missile control system

The additional feedback loop increases the speed at which the missile responds to an error, The fundamental automatic control principle of using the reaction of the controlled item as a basis for further and continuous correction has already been mentioned. In a missile system the controlled item is missile attitude. If this reaction were the *only* guide to further correction, it would occur too late to provide fine control. The reaction would be late due to aerodynamic lag. When the control surface is deflected, it takes an additional period of time before the missile returns to the desired attitude. The deflection of the control surface determines the force tending to move the airframe, not the airframe position directly.

Without a followup loop, a missile would tend to oscillate about the desired heading. Suppose a missile without followup turns to the right due to a gust of wind. The sensor detects the error and deflects the rudder to the extreme left to get the missile back on course. A left rudder is maintained as long as this error is detected. After a short time the missile is returned to the desired heading. The sensor detects this fact by means of the major feedback loop, but it is too late, since the swinging missile is now well on its way toward a left deviation which requires a right rudder. The operation repeats itself, creating cyclic movement dependent on the missile aerodynamics and the control-system characteristics.

The followup loop sends back earlier information since it does not include the lag of missile reaction. The signal is not without delay, however, since the control equipment is not instantaneous.

This system lag is shown in the figure on the right which combines error and followup signals. Consider any instant during the period that the error signal is increasing. At this instant the followup signal is less than the error signal. This produces a resultant signal. A certain lag time is required before the followup signal reaches the error signal amplitude which existed at the instant considered. This time is shown as "lag" in the figure.

The figure shows that the followup signal is added to the error signal with such polarity (or phase, if the signal is modulated AC) as to *oppose* the error signal. This type of feedback is therefore *degenerative*. It can also be called inverse or negative feedback. It is comparable to the inverse feedback found in many electronic amplifiers.

The signals are combined to produce a resultant which is the output from the mixer circuit. Mixers were described in Section II of this chapter. The resultant is the corrector signal which operates the control surface.

FOLLOWUP GENERATORS DETECT SYSTEM OUTPUT

The devices which detect system output are called followup generators or control-surface pickoffs. They are a position-detecting device such as a selsyn generator or a resistance pickoff.

Followup generators are mechanically connected to the actuator, control surface, or the linkages between. Similarly, a followup generator can be connected to jet vanes or movable steering jets if they are used to provide missile correction. Every generator is nulled at the streamline position of the control surface.

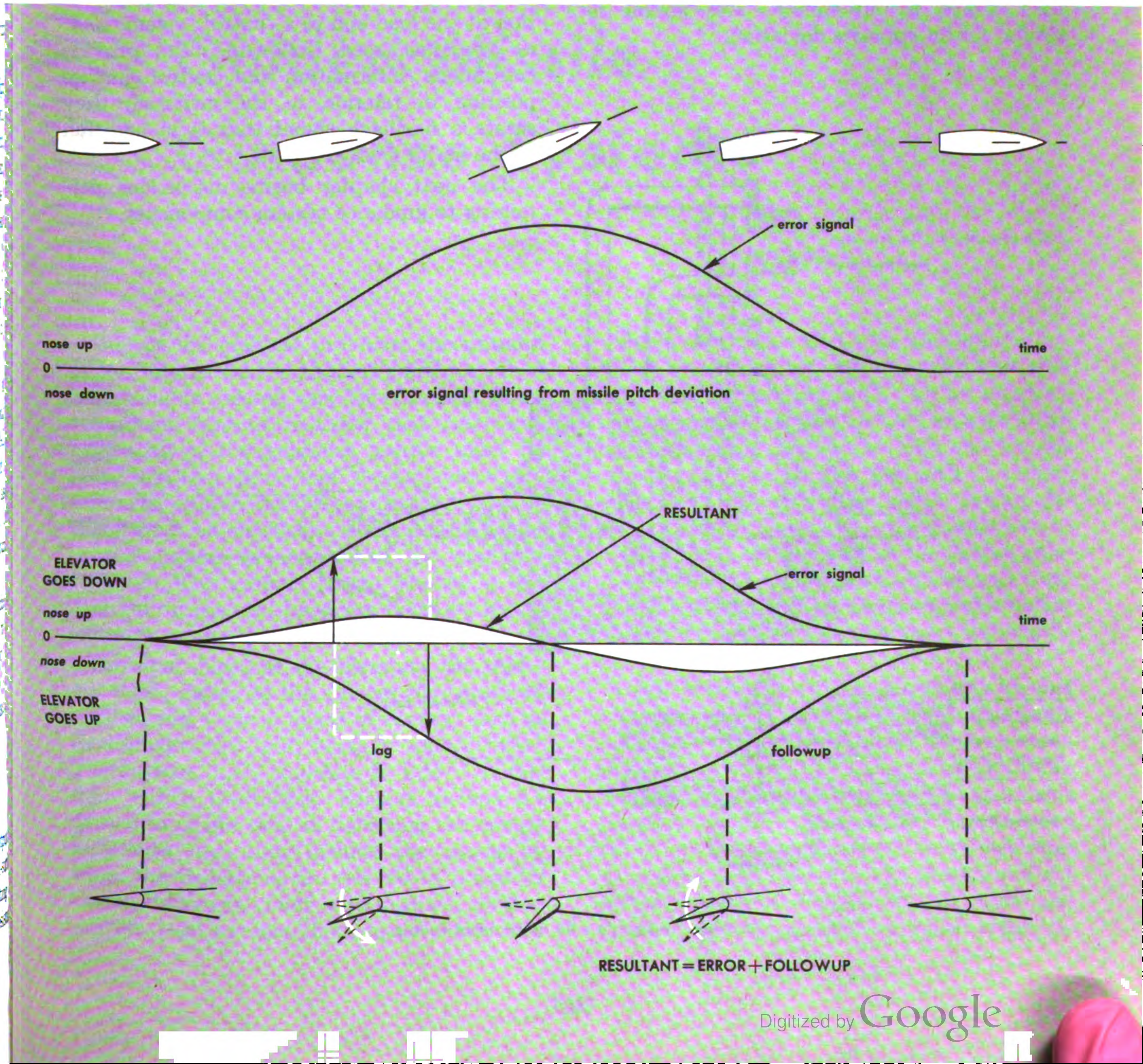
The selsyn generator and resistance pickoff are similar to the pickoffs that are used with gyros to detect missile attitude. These devices are not covered here since they were covered in the section on pickoffs.

FOLLOWUP SIGNALS CAN BE TRANSMITTED MECHANICALLY

One type of followup path exists in which the information is transmitted mechanically rather than electrically. The control surface position is indicated by a force which is imparted back to the controller by means of a spring. The action of this spring is shown in the illustration of the next page.

Suppose an air-pressure signal from a pneumatic pickoff moves the diaphragms of the air relay upwards. The followup arm then moves clockwise. This mechanically moves

Relationship of followup to error signal



the valve spool of the air valve upwards. High-pressure air then forces the piston of the air actuator to the left. This compresses the followup spring and tends to rotate the followup arm *counterclockwise*. Since the followup force opposes the original motion of the followup arm, the feedback is inverse. This is as it should be.

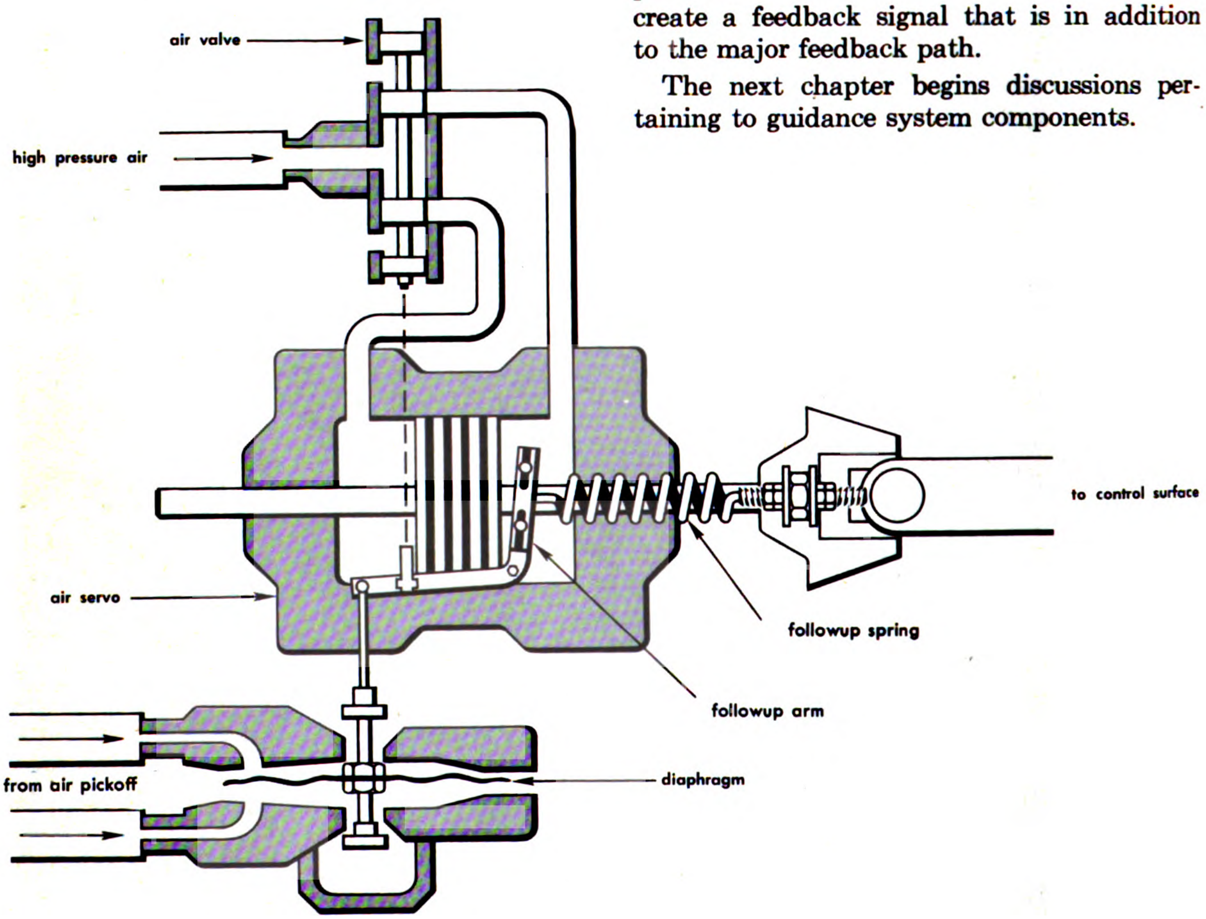
A large signal will create a larger surface deflection before the feedback force is great enough to return the followup arm back to zero. The spring will then force the followup arm and air valve in the other direction to move the surface back. Therefore, the spring acts to limit surface deflection to a value dependent on the error signal and to return the surface to streamline. The action is similar for an opposite deviation.

FEEDBACK LOOPS COMPLETE CLOSED CYCLE

This discussion of feedback loops brings to an end our coverage of the units that make-up the closed cycle of a missile control system.

Sensor units, we found, detect missile deviations from desired flight conditions. Computer units change signals to represent additional information. Without reference units, a control system would not be able to make desired changes in flight conditions. The fundamental purpose of amplifier units is to produce an enlarged reproduction of the input signal. Controller units in missile control systems control the operation of actuators. Actuator units accomplish energy transfer and transformation at the load end of a control system. And, as you just read, followup units provide continuous information on control surface position in relation to the airframe; they create a feedback signal that is in addition to the major feedback path.

The next chapter begins discussions pertaining to guidance system components.



Air relay and servo with mechanical followup

Chapter 6

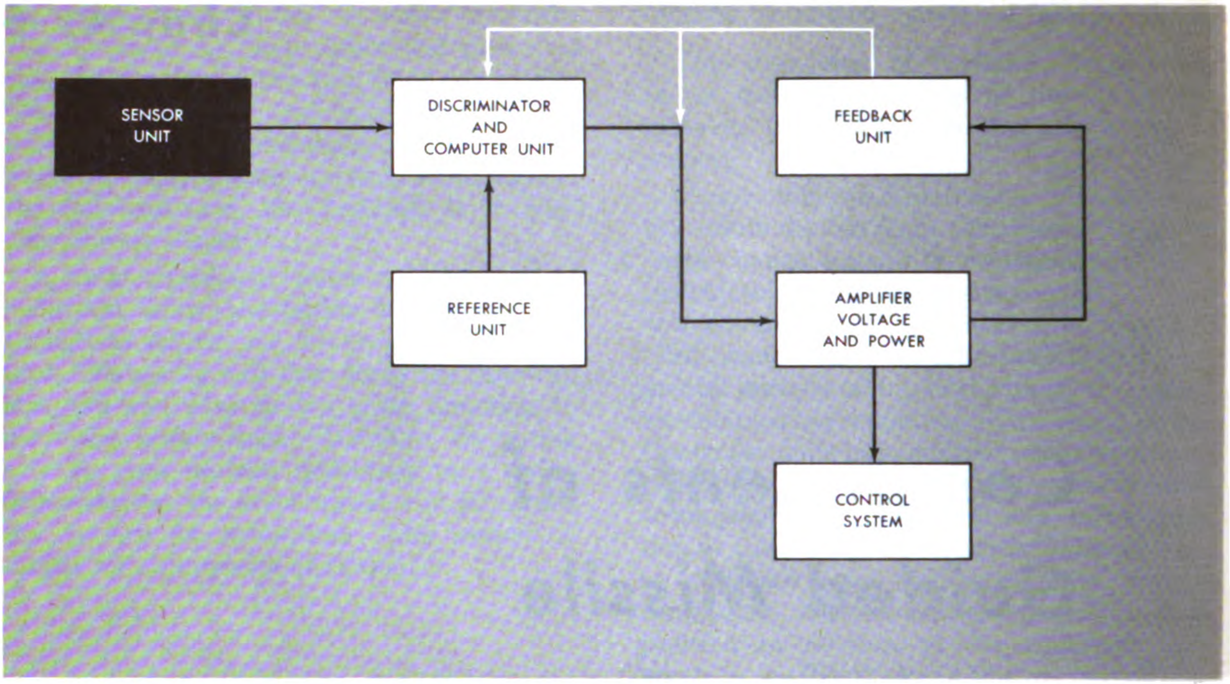
Components of Guided Missile Guidance Systems

One of the greatest problems encountered in the missile field is that of guidance. It is impossible to preceed very far with missile development without realizing that the characteristics of the guidance system must be given first-order attention. Any guidance system must employ certain components; that is, there must be an adequate number of stages in the system to accomplish the guidance function. The various stages, shown in the diagram on the next page, are taken up in the following sections.

The first block of the diagram, the *sensor unit block*, refers to the devices that are used in detecting various forms of energy, such as electromagnetic, light, heat, or sound waves. The devices then translate it into a usable form within the missile. Sensors include the devices that are used in detecting the energy radiations, and they include the means by which the radiations are relayed to the other circuit elements. They also include accelerometers used to sense changes in motion.

From the sensor unit, the information signal is passed to the *discriminator and computer unit*, the next block in the diagram. Here the signal is "interpreted" so that it can be applied to circuits that will actuate the control surfaces. Here we find decoders, integrators, differentiators, signal mixing devices, and the computer. This unit can be considered as the "brains" of the missile guidance system. It makes calculations and comparisons of external and internal imformation at microsecond intervals to keep the missile on course.

Also feeding into the discriminator and computer unit is the *reference unit*. To obtain a set of guidance conditions, it is often necessary to compare two sources of information. This information may come from an outside source or it may be some form of recorded information that was put into the missile prior to launch. It is, therefore, the function of the reference unit to establish the conditions and refer them to the other circuitry for computa-



Stages of a missile guidance system

tion, amplification, and power conversion to the control system.

The guidance signals, after proper recognition, must be amplified sufficiently to become usable in the remaining units of the missile as represented in the block diagram. The *amplifying block* contains circuit elements that give "energy" to the "muscles" of the system. These amplifying devices may react to voltage, current, frequency, or combinations of these and, then, increase their operational effect.

The "muscles" of the guidance system is the *control system block*. Relays, solenoids, servo motors, and other actuating units go to work, putting into operation the result of numerous calculations, comparisons, and amplifications of the initial guidance request made to the system.

The *feedback unit block*, though it may not contain any specific circuits to accomplish its purpose, must be taken up as a definite member of the block diagram. Feedback can be thought of as that part of a guidance intelligence system which helps the system to distinguish right from wrong. It is constantly observing the guidance signals and making attempts to increase or decrease their effect so that there will be a smooth acting system. If any unwanted signals enter the system, it also tries to eliminate these by feeding back information signals to the previous circuits.

It should be understood that the block diagram is not only related to the functions that take place within the missile itself but also to various outside factors as well. With this in mind, you will later be able to see how the reference unit may refer to a ground base unit that is setting up the reference. The computer section may or may not be contained within the missile, or the section could be a combination of both outside equipment and missile-contained equipment that is tied into the operation of a missile.

The form of guidance block diagrams may vary, but the final results of the systems described by them would be the same. As the study of this guidance loop is made, it should be kept in mind that new methods are constantly under development. It is advisable that frequent reference be made to current progress reports in order to keep up with new systems.

Sensor Units for Guidance Systems

In this section we will consider the first block of the guidance system diagram, the *sensor unit block*. As stated before, sensor units include devices that detect various forms of energy, such as electromagnetic, heat, light, or sound energy.

Guidance can be performed by any one of several different methods. *Methods*, as used here, refer to the overall picture of guidance procedures. There are many systems that could be used to accomplish any of the methods. Some of the systems are taken up in relation to the sensing element that is used in conjunction with them. Below are listed some of the general methods of guidance:

a. Preset

b. Command

c. Beam rider (direction along a beam)

d. Hyperbolic (direction by navigational network)

e. Celestial navigation

f. Terrestrial reference (including earth phenomena)

g. Homing

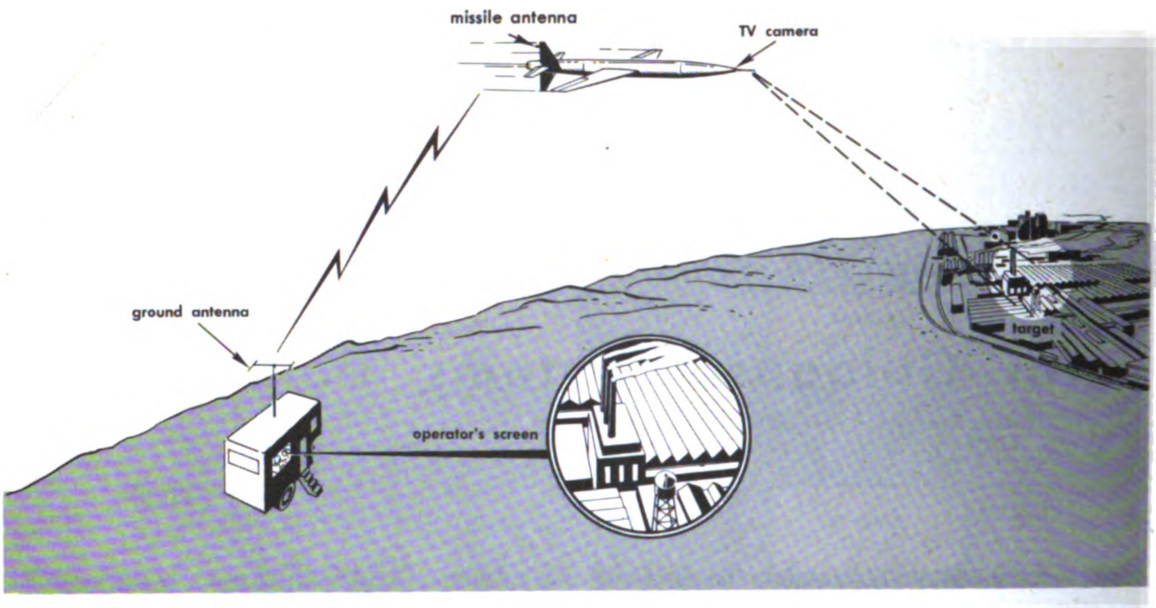
METHODS USED IN MISSILE GUIDANCE

To enable you to understand and appreciate the problems that may arise in choosing a specific guidance system, it is well to first have an overall picture of the guidance methods listed above.

In the preset method of guidance, the control equipment is contained wholly within the missile. All adjustments as to course, altitude, speed, and dump-point must be made prior to the launch time. The idea of having a missile fly a predetermined path in such a manner has both advantages and disadvantages. A major advantage of such a method is that countermeasures against such a missile are quite limited; a major disadvantage is that after once launching the missile, no correction can be made for factors not previously anticipated.

In the *command* method of guidance, the intelligence for the missile's guidance comes from outside the missile. The missile contains a receiver that is capable of receiving directions from a ground station or mother aircraft and executing these commands through the control system and the control surfaces. Usually, in the command system, several channels are operated from the control point by modulating the transmitted signal at several frequencies. Therefore, changes in altitude, direction, speed, or any other factor may be made as desired by the controlling operator.

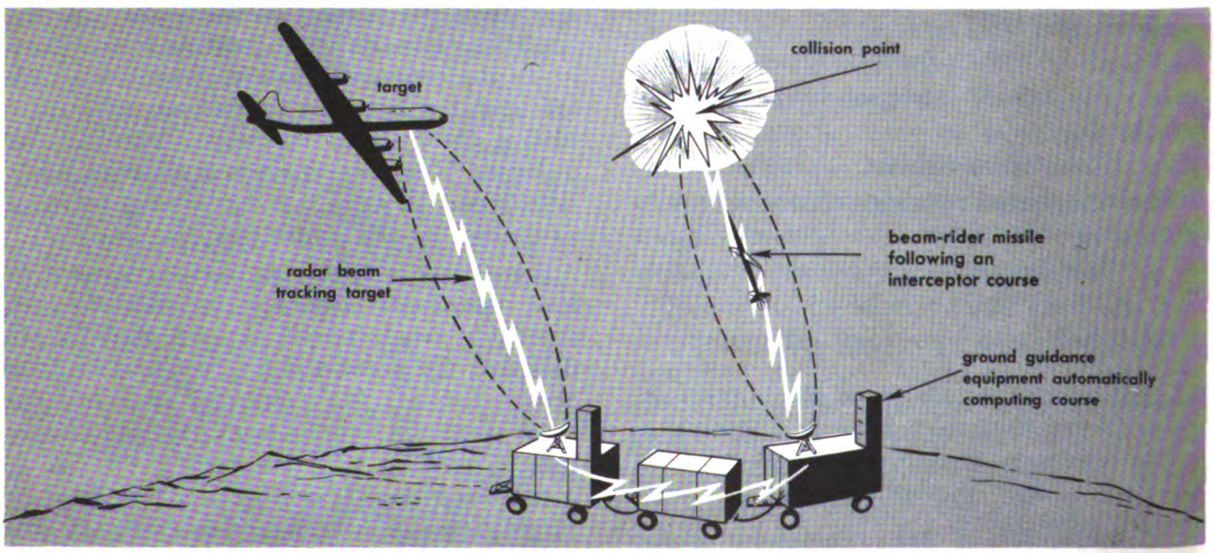
Sometimes a television system is used to give additional accuracy to this method, especially when the mother aircraft is endangered by accompanying the missile over hostile terrain. The controller then has the same viewpoint as if he were physically in the missile, guiding it to the target. The use of a television repeat-back system is limited to target areas that are not overcast. A disadvantage to the command system is that it is generally subject to enemy jamming.



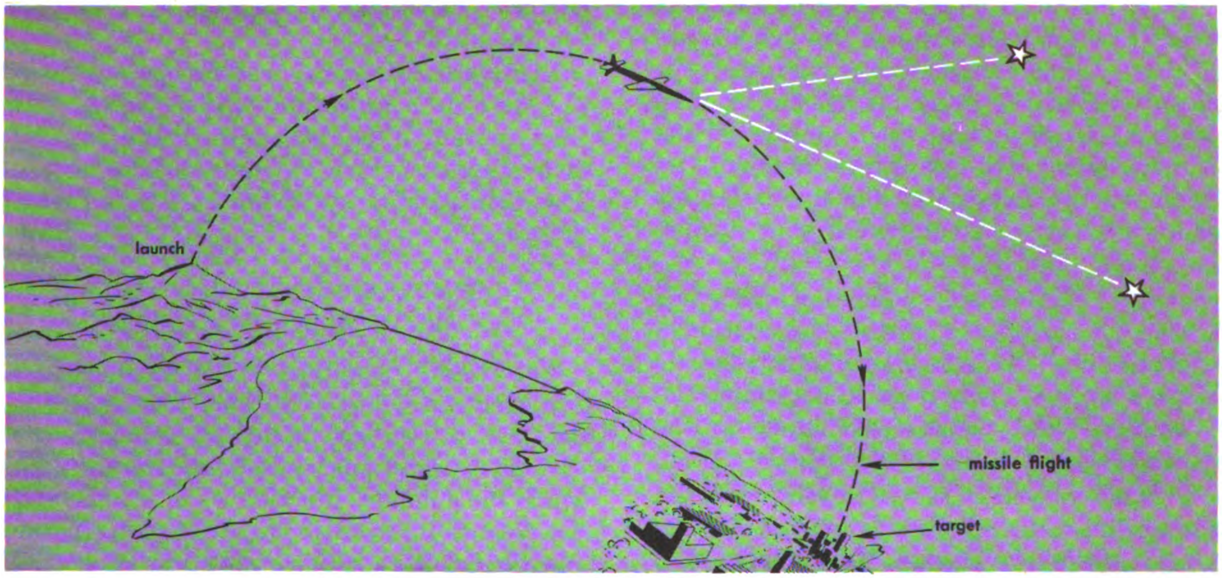
Television provides a means of guiding a missile to a target

When using the *beam-rider* method, a missile contains equipment which enables it to follow an electronic beam. In this method, the missile is launched into the beam. As the beam is directed toward the target, the missile tends to keep itself centered on the beam until it explodes at the target. This beam should not be confused with the radar beam that follows (tracks) the target. The two beams are illustrated below.

The *hyperbolic* method can be used as a long- or short-range navigational network. Such a system consists of master and slave stations putting out low-frequency pulses at constant intervals. The slave station is triggered by the master station and transmits its pulse a few microseconds after the master pulse. By using automatic computers contained in the missile, the missile establishes position by an imaginary line of position set up by the master and slave stations.



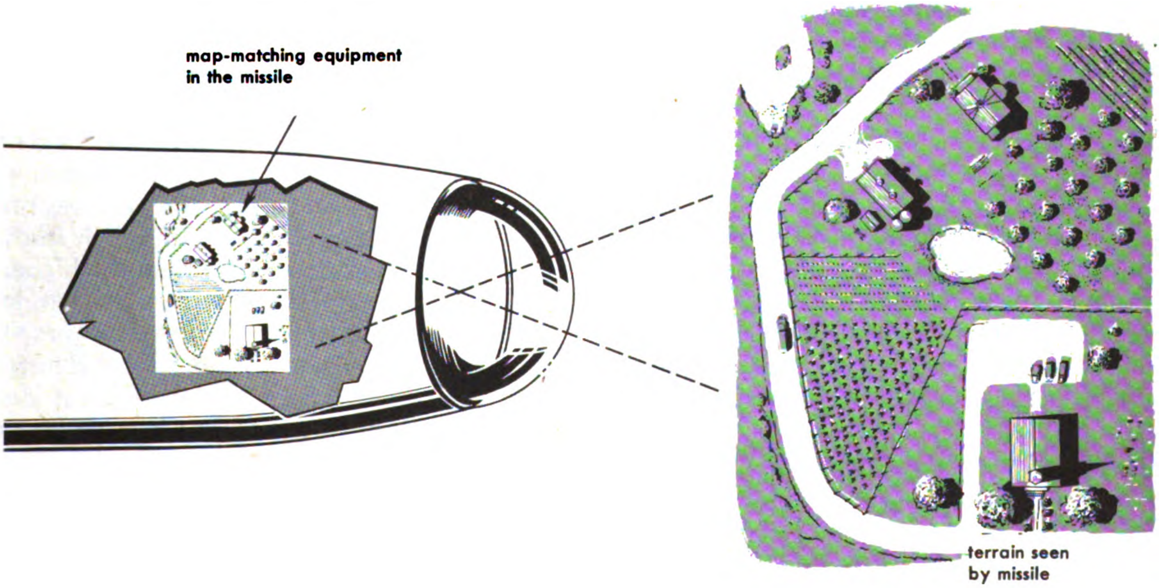
Beam-rider method of missile guidance



Celestial navigation can be used to guide long-range missiles

The *celestial navigation* method is highly complex. It consists of a mechanism that takes celestial fixes and, through electronic means, keeps the missile on course. This will be used primarily for long-range guidance Its general operation is somewhat as follows:

Star-tracking telescopes take a fix on some predetermined star, and information thus obtained is fed to an electronic computing device that determines the position of the missile relative to the earth. Necessary controls are activated by the error voltages developed if the missile does not coincide with predetermined values set into the missile. It is fairly safe to assume that this system would be used on long-range missiles since it is necessary for the missile to fly at high altitudes above the weather where the stars are constantly visible. This method is illustrated in the figure above.



Map-matching by electronics equipment determines missile position

In the method involving *terrestrial reference*, missile position is determined through the use of charts, some characteristic property, or phenomena of the earth. A system of map matching is commonly used. A comparison of maps that were previously obtained by reconnaissance missiles to the target area is made by electronics equipment contained in the missile.

Systems are also being devised which guide missiles by using the earth's gravitational, magnetic, and electrical fields. Much cosmic research work is being done to determine the effect that these cosmic rays have on the earth's magnetic and electrical fields.

A *homing* system is employed, in some manner, on most missile missions. A homing system serves its purpose at the end of a flight when the guidance system must distinguish the target from the target background. Among the ways available to guide missiles into the targets are radar, heat, and light homing devices.

Radar homing systems are generally divided into two types. In one type a radar transmitter at a ground location or in a mother aircraft directs a beam toward the target; the reflected pulse of energy is detected and measured by the radar receiver in the missile. A highly directional antenna system is used in both the transmitter and the receiver installation. This system's main disadvantage is that the transmitting equipment needs to be within line-of-sight distance from the target.

In the other type of radar homing, the missile itself transmits and receives the radar pulses as illustrated to the right above. Here again, the target is "illuminated" by the radar transmitter, thus giving the missile a target that stands out well from the target background. This system of homing is well adapted to both short- and long-range missiles. Radar homing systems are subject to enemy jamming.

Heat-homing devices depend on the fact that radiations are given off from bodies at intensities that differ with those of their surroundings. These radiations, or infrared rays, are detected by a device that translates these variations into voltage changes. The voltage changes are amplified into steering directions for the missile.

Light-seeking homing systems operate much the same as heat-seeking types, except that they contain a photoelectric device instead of a thermal device to home on the target. The light-seeking types are limited by visibility conditions.

SENSING ELEMENTS EMPLOYED IN MISSILE GUIDANCE SYSTEMS

The overall operation of the methods just discussed are taken up in later chapters. For the present we are going to consider the various types of energy, or phenomena, with which the guidance function is performed, and the devices that detect these energies. In the selection of a type of sensing unit that will detect the forms of energy and guide the missile to its target, consideration must be given to many items, some of which are:

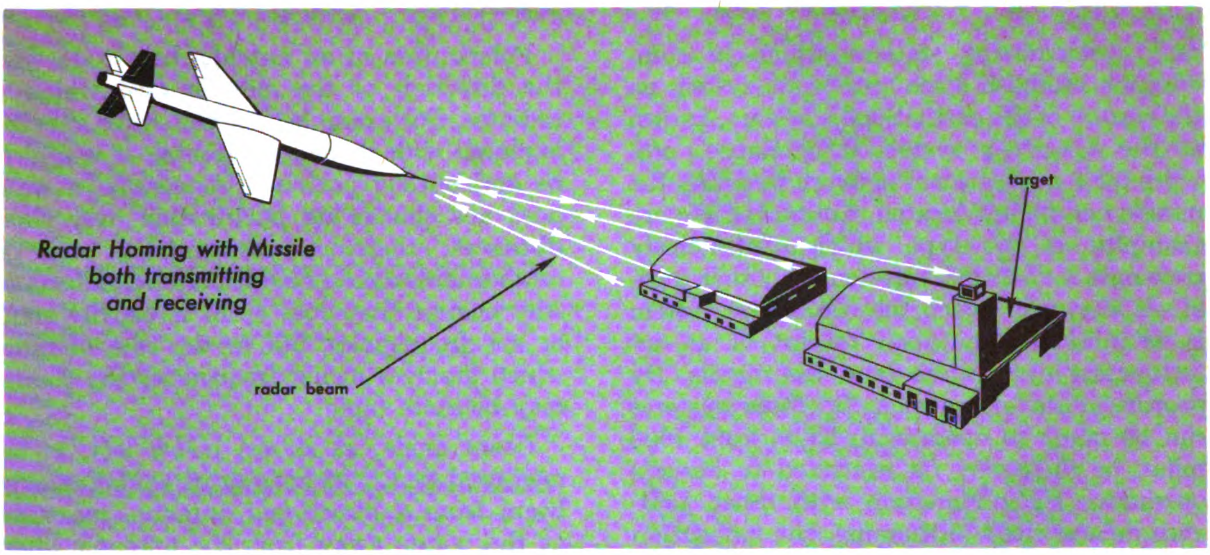
a. Maximum range.
b. Information required.
c. Accuracy.
d. Operating conditions.
e. Viewing angle of sensor.
f. Sensor size and weight.
g. Type of target.
h. Relative speed of target.

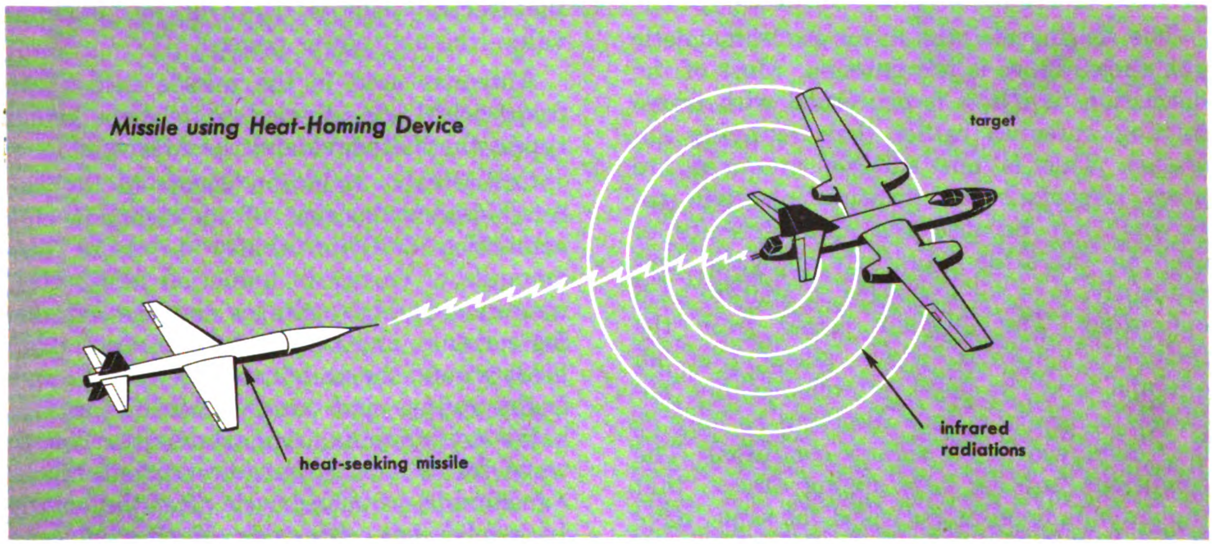
Sensors Using Electromagnetic Waves

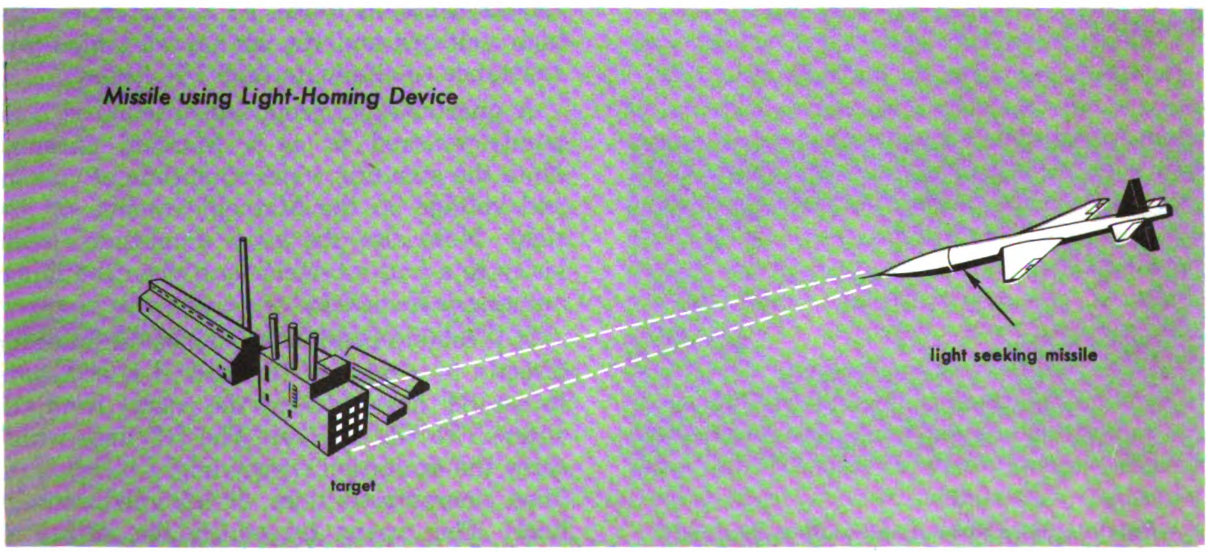
Included within the electromagnetic methods of guidance are the systems operating by means of radio or radar transmission.

RADIO SENSORS OF MISSILE GUIDANCE SYSTEMS. Use of a radio system, or beacon, in guidance units depends on suitable transmission into the target area. Although radio systems have a greater range than the higher-frequency radar systems, they are not yet dependable unless operated under proper conditions. The effects of atmosphere changes limit the use of this system over great distances. Higher-frequency electromagnetic waves are little affected by weather, but are limited to line-of-sight distances since they are not bounced by the ionosphere.

Guidance by the use of radio can be accomplished in a number of ways. Most of these ways depend on the use of two or more

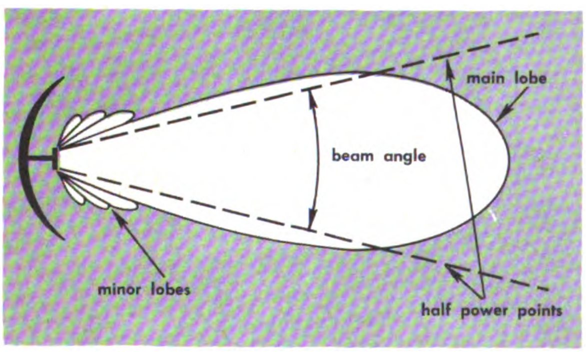

Radar Homing with Missile both transmitting and receiving
target
radar beam


Missile using Heat-Homing Device
target
heat-seeking missile
infrared radiations


Missile using Light-Homing Device
light seeking missile
target

radio transmitting stations which in combination establish some coordinate system on the surface of the earth. Such a system, known as the *hyperbolic* system of guidance, provides for a navigational network in which the missile can be self-located. *Loran* and *gee* are examples of the type which continuously transmits a coordinate system. Another variety is *shoran* in which the coordinate system is established only upon receipt of an interrogating radio wave from the missile. By proper choice of wavelength, these systems can be made to operate up to at least 1000 miles, or they may be restricted to horizon operation. In combination with other guidance devices used for homing, the navigation network is one of the most versatile types.

The sensing element in an electromagnetic guidance system is the antenna. Most of the discussion to follow is concerned with transmitting antennas rather than receiving antennas. This is not because antennas are more important for transmitting than receiving, but because the characteristics of each are the same if properly analyzed; consequently, it is not necessary to analyze both operations. This likeness comes from the reciprocity theorem which states that an antenna may be used interchangeably between transmitting and receiving as long as it is employed under similar conditions. The conditions imposed are, among others, that the antenna must be polarized the same, have the same length, and operate on the same frequency.



Beam angle is the measured width in degrees of an R-F energy beam

Polarization of an antenna has to do with the direction of the electric field which it radiates. For example, if a half-wave antenna is placed with its axis horizontal, it is said to be horizontally polarized. Generally, a wave does not change its polarization over short distances; therefore, transmitting and receiving antennas should be oriented alike. Any change that does occur is usually small at low frequencies, but in the case of radar which involves higher frequencies the change may be quite rapid.

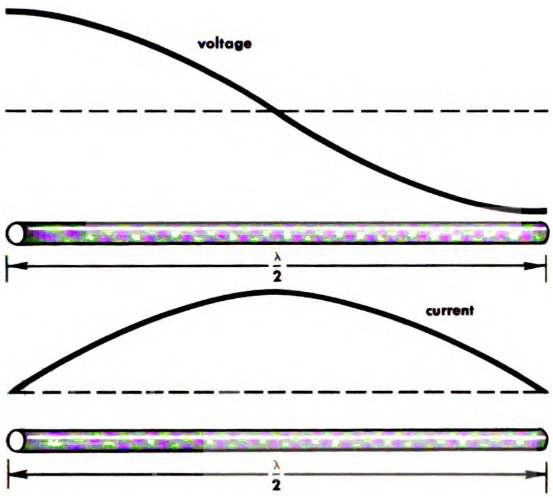
Beam angle, as shown below left, is the measured width in degrees of an RF energy beam from a directional antenna, measured at the points away from the antenna where the electric field is 0.707 of its maximum value (the half-power points).

Antennas used in the missile field at RF frequencies are divided into two general types: the half-wave and the Marconi.

Considerable radiation occurs in the *half-wave* element because of its resonant characteristics and its ability to store large amounts of energy in induction fields. Resonance causes high voltages and high circulating currents, and they in turn produce strong fields around the antenna. Shown on the right is the voltage and current distribution for a half-wave antenna. An examination of the wave forms shows that the current reaches a maximum a quarter cycle after the voltage. In application, the ends of the half-wave (Hertz) antenna must be insulated for the high voltages and the center must have low resistance to eliminate high I^2R losses.

A quarter-wave grounded antenna is commonly referred to as a Marconi antenna. Note the standing wave amplitude of current and voltage on the Marconi antenna on the right. Note also the wave amplitude's similarity to a half-wave element when the image antenna is included.

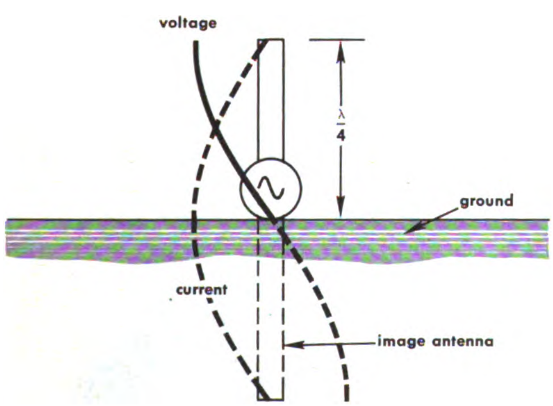
It is seldom possible to connect an antenna directly to a receiver. Instead, it is necessary to use RF lines to carry the energy from the antenna to the receiver. The RF lines which carry the excitation energy might be resonant lines, nonresonant lines, or a combination of both. The means by which the energy is taken from the sensing element is referred to



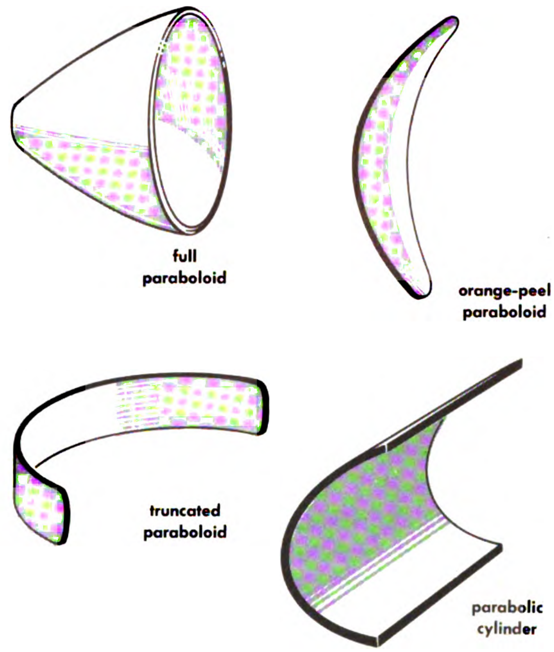
Voltage and current distribution for a half-wave antenna

as the *sensing element pickoff*. Pickoffs include any of the devices that are used to transfer the energy received at the sensor to the following detecting and amplifying stages. These devices are taken up later in this chapter.

RADAR SENSORS OF MISSILE GUIDANCE SYSTEMS. Present radar systems are intended primarily for air or ship targets, with some application to land-based targets that afford a good target-to-background signal. A main drawback of this system is its limited range. It is, however, practically unaffected by weather conditions; it can be operated day or night and it can be adjusted for minimum interference by jamming transmitters.



Standing wave amplitude of current and voltage on Marconi antenna



Parabolic reflectors

At higher frequencies, the primary functions of radar antennas are to concentrate most of the transmitted energy in one direction and to give a usable directional pattern in the same direction. The directional features of radar antennas are usually their most important characteristics. A directional antenna system makes it possible to illuminate a specific target area and receive the reflected pulses in a selected direction. Directivity of both the transmitting and receiving antenna improves the overall efficiency of the system, thus minimizing enemy countermeasures.

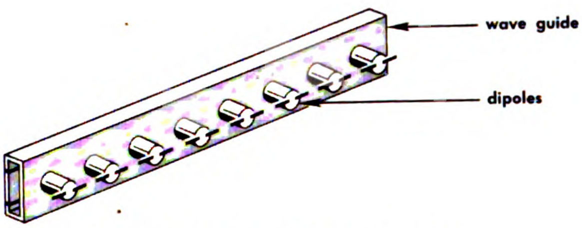
The basic method of obtaining directivity in an antenna system is to space two or more simple half-wave elements so that the fields from the elements add in some directions and cancel in others. A set of antenna elements is called an *antenna array*. Among the common types of arrays are the broadside, end-fire, collinear, and the parasitic or Yagi array.

Another simple method by which directivity of an antenna system can be obtained is with the use of parabolic reflectors. In this type of antenna system, the driven element is placed at the focal point of a parabolic reflector.

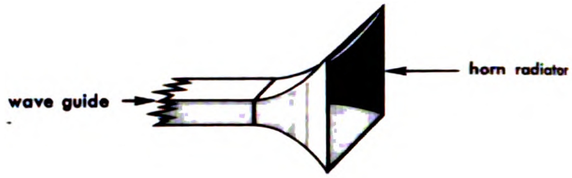
The antenna is then working into a constant-phase reflecting surface. The pattern of

field intensity is altered by changing the focal length of the paraboloid, distorting the reflector, or varying the field intensity. Four types of parabolic reflectors are shown on the preceding page.

Another means of concentrating power is by wave reinforcement. This concentrating of power is accomplished by one or more rows of dipole radiators, spaced some distance apart, radiating at a fixed phase with respect to each other. Dipoles may be arranged along the length of a supporting rod, as shown below and excited to yield a narrow endfire at the end of the rod, or the dipole arrangement may give a nondirectional pattern in the horizontal plane and have a narrow pattern in the vertical plane.



Collinear array of dipole radiators

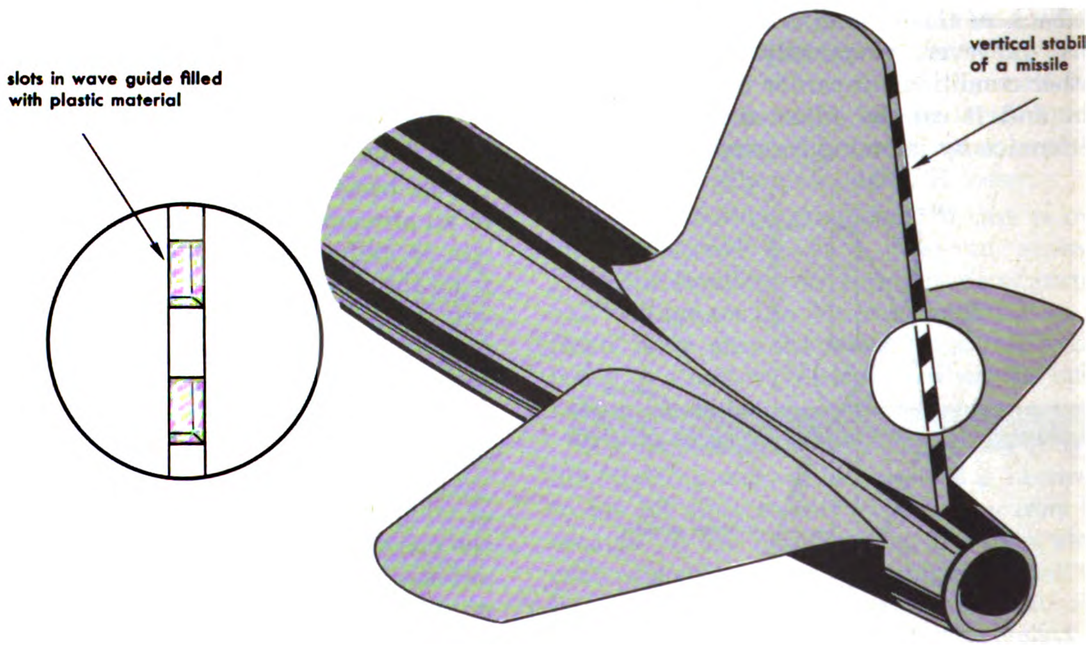


Horn-type radiators are used with wave guide type feeds

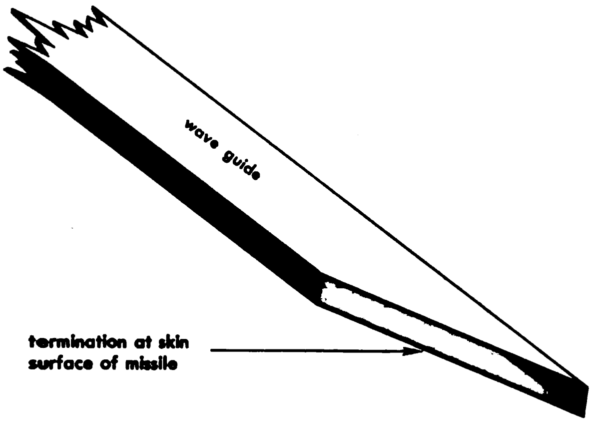
When slots of the proper size and shape are cut into a waveguide, they act as radiators the same as a dipole antenna. The slot arrangements are generally found in higher-frequency radar systems. Collinear slot arrays are used in place of dipole radiators in order to obtain a smoother aerodynamic surface, as illustrated below.

Energy is also beamed by the use of horn-type radiators, all points of the wave front being in phase at the mouth of the horn. The horn is generally used in place of the dipole and reflector arrangement. In general, horns are best adapted for use with wave guide-type feeds, while dipoles are used with transmission line feeds.

As higher speeds are attained in the missile field, new antenna units must be developed to eliminate drag. For this reason, slotted, wave guide-termination, and recessed antennas are being developed.



Collinear slot arrays permit smoother aerodynamic surfaces



Zero termination antenna

Mention already has been made of the *slot* type. Energy can also be radiated from the wave guide itself if it is properly terminated. The beam angle and radiation pattern is governed mainly by the angle at which the wave guide is terminated. *These wave guide-termination* types of antennas, pictured above, are generally called zero termination.

If a radiating stub is made to lie within a surface rather than protrude, it can radiate a pattern much the same as a protruding type. Such a *recessed antenna* is illustrated below. Recessed antennas are invaluable for missiles that are designed to travel at high Mach numbers. They offer little wind resistance and can be located more easily on a missile body.

The antennas just discussed are only a sample of the numerous types that could be used as the electromagnetic sensing elements of missiles.

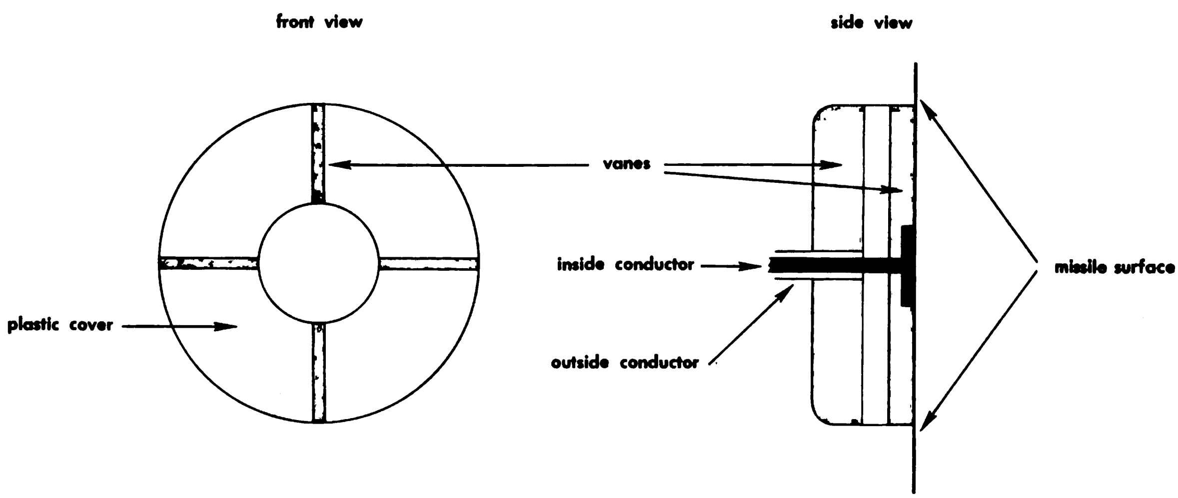
For the purpose of guidance, there are three general types of radar systems that are used. These are:

a. Pulsed
b. Continuous-wave (CW)
c. Frequency-modulated (FM).

Important factors that are considered when choosing equipment are the relationships between wavelengths of transmission, size of antenna, maximum range, and angular discrimination of each. Generally, FM radars operate on longer wavelengths than pulsed and CW radars; consequently, FM radars require large antenna arrays to obtain the narrow beam needed to give good angular discrimination.

Angular discrimination and maximum range increases as frequency increases; antenna size decreases with wavelength if angular discrimination, maximum range, and other factors remain constant. The higher-frequency systems are usually not desirable because technical difficulties and echoes from clouds, cold fronts, and rainstorms increase with frequency.

Pulsed systems provide convenient means of target selection in that they can be gated in such a manner that the echo from only



Recessed antenna

Type	Maximum Range	Angular Discrimination	Range Accuracy	Range Discrimination	Antijamming Features	Special Weaknesses
Pulsed	Good	Good	Good	Good	Good	Poor Minimum Range
FM	Poor	Poor	Excellent	No Apparent Usable Feature	Poor	Short Range
CW	Good	Good	No Range Data	None	Poor	No Target Discrimination

one target will be used to guide the missile. A pulsed system also decreases the possibilities of jamming due to the gating principle.

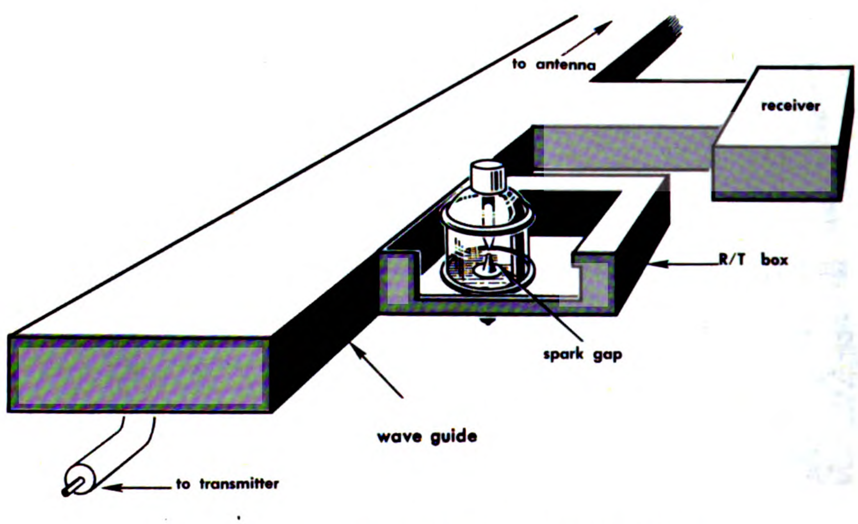
The above table gives a comparison of the three systems mentioned. It can be seen by these comparisons that the pulsed radar is the best choice except for range characteristics.

In RF systems in which one antenna is used to transmit and receive, the switchover is usually made mechanically. Such a method is not used in radar systems in which switching is accomplished automatically. A transmit-receive (TR) switching device functions automatically. It may be in the form of resonant lines, T-joints, or spark gaps. It is the function of these devices to permit transmission of a signal from a single antenna system without subjecting the receiver to damage by the high power being transmitted. It is also necessary to "cut off" transmission at the proper time in order to receive the reflected signal.

Illustrated below is a typical spark-gap R/T box. Upon transmission a spark is created. The spark causes an open circuit to be reflected at the wave guide leading to the receiver. Therefore, none of the high-power energy is allowed to enter the receiver. After transmission, the spark gap is extinguished; the physical dimensions of the wave guide then reflect a short circuit in that portion of the wave guide leading to the transmitter. This reflection causes the incoming signal to be diverted into the receiver section.

Infrared Sensors

It is necessary to study the intensity and wavelengths of the infrared radiations of a



Spark-gap type of transmit-receive switching device

target in order to successfully employ guidance systems which operate on the principle of receiving intelligence by interpreting the infrared radiations of the target. Many targets emit a sufficient amount of infrared energy to permit their detection without using an outside means of illumination.

Most of the infrared seekers are comparatively slow in reaction time and quite limited as to distance of operation. Fog and rain limit their effective operation, and decoy fires may cause false seeking. Considering the altitudes that some missiles attain, the heat-seeking device may not be sensitive enough to discriminate between target and background when the pre-dump point is reached. With increasing sensitivity and reaction time, infrared seekers are being used in various surface-to-air missiles.

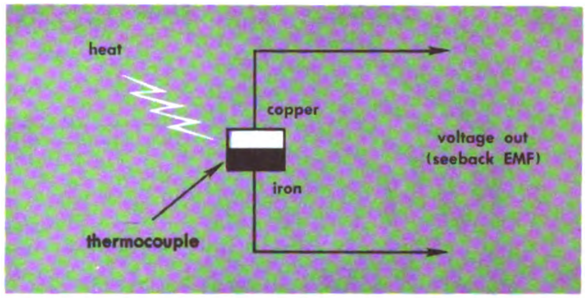
Since heat-seeking devices depend on the infrared radiation given off by a target, the detection of infrared radiation has a definite advantage in that it permits the construction of a simple detecting device and does not require a transmitter.

Infrared radiation of energy in a spectrum is the radiation that is invisible to the eye. Any object that is at a temperature different from its surroundings radiates heat in various quantities. This heat energy applies to the frequencies between 7000 and 3,000,000 angstroms.

There are two methods of detecting a target by means of infrared radiation. *Characteristic-radiation infrared detection* is the term used to describe the process of detecting emitted infrared radiation which is characteristic of the target. This method is considered as a *passive* method. Targets are also detected by an *active* method. This active method operates on the infrared radiation that is detected when a beam of infrared is sent out by a special transmitter and then reflected by the target.

The passive method of infrared detection is far more important and advantageous than the active method because the former permits the use of a simple detecting device and does not need a transmitter.

Radiations of infrared energy can be detected with a variety of sensing elements.



Thermocouple

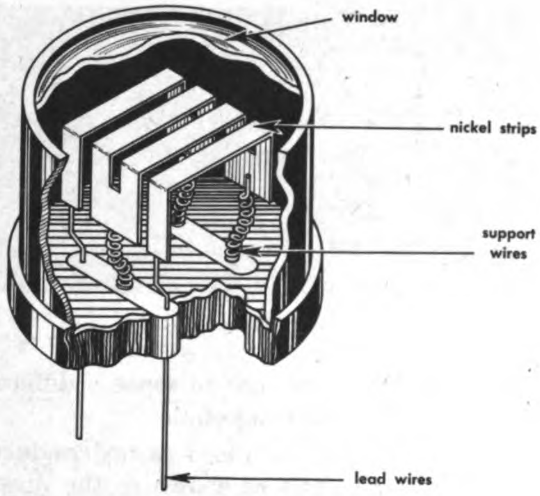
One of the first devices to sense a difference in heat was the thermocouple.

A *thermocouple* is a joined pair of conductors of dissimilar metals as shown in the illustration above. Any change of temperature at the junction causes a measurable voltage difference to exist between the two. This voltage is known as the *seeback* electromotive force.

Although the voltage difference at the junction of a thermocouple is quite small, the sensitivity can be increased to a point where it becomes a useful detector. This increase is accomplished by utilizing a number of such devices in series with an appropriate circuit. When several thermocouples are joined in series they are referred to as a *thermopile*. A thermopile detector may be mounted at the focal point of a parabolic reflector to increase the intensity of the radiation and to provide a means of sensing the direction of the radiation.

A *bolometer* is another type of sensing element. It is a heat-sensing element which depends on the change of electrical resistance of a material for its action. The first bolometer was made in 1880. It was made of two thin strips of platinum, which formed the two arms of a Wheatstone bridge. The strips were blackened on one side. As heat was applied to the strips, the absorption of energy by one of the strips caused an increase in its electrical resistance. This change of resistance causes a change of current flow when an external voltage was applied to the circuit.

The most popular type of bolometer in use today is the nickle-strip bolometer illustrated on the next page. This bolometer consists of four nickel strips which are supported by springs of phosphor bronze. These springs

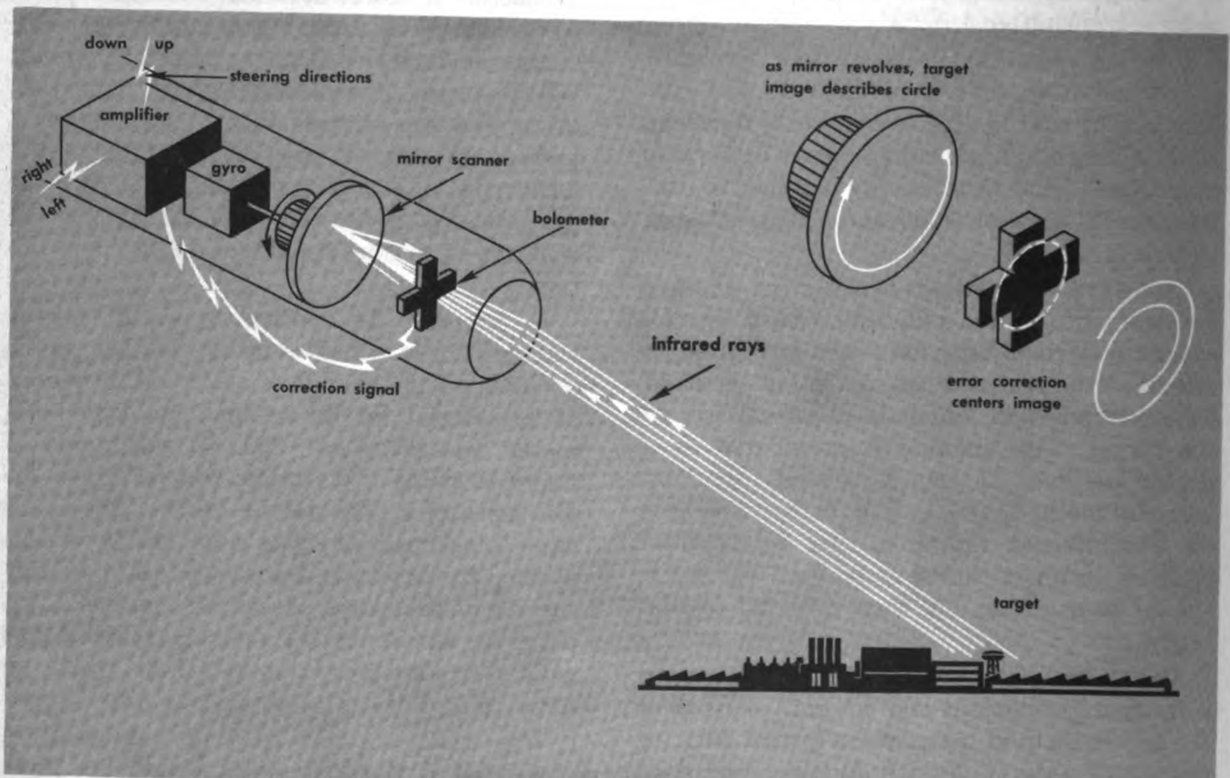


Commonly used bolometer

are supported on a mounting to which lead-in wires are attached. A mirror focuses the heat signal from the target on the bolometer. Note the use of a bolometer in the illustration below.

The requirements of thermosensitive elements are high sensitivity, freedom from microphonics, and low time constant. Thus, various thermosensitive surfaces are being experimented with to give desired results. The lead-sulfide (Galena) cell is one such heat-sensitive element that shows great promise for infrared-seeker usage.

The typical operation of a bolometer-type infrared system may be somewhat as follows: An *eye* — consisting of the bolometer, a front covering window, and a commutator — scans or moves over the target. Radiations from the target are reflected onto the bolometer face by the scanning mirror. When there is no radiation entering the eye, the bridge circuit that notes the change in resistance of the bolometer is in balance. When radiations are present, this causes an unbalance of the bridge; this unbalance causes a signal to exist at the grid of an amplifier. The commutator picks off the changes in resistance as the mirror scans the field of view. It is in synchronization with the mirror and closes



Missile guidance system using a bolometer

appropriate switches so that when the mirror sees the upper part of the scanned circle the *up channel* is closed; when the mirror sees the right side of the scanned circle, the *right channel* switch is closed, and so on.

Light-Sensitive Sensors

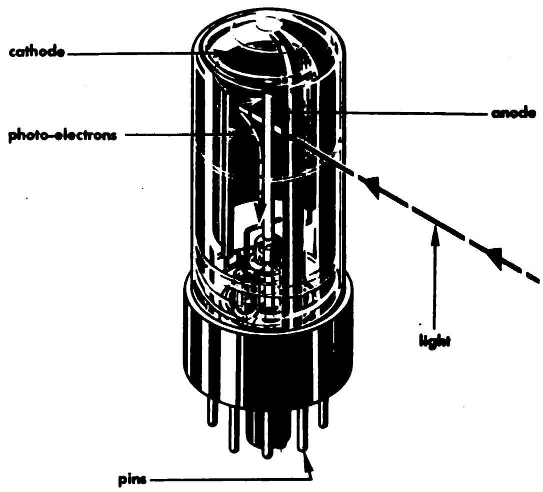
Hertz discovered, at the end of the 19th century, that electrons were ejected from certain metallic surfaces when exposed to light. As a beam of light strikes such metallic surfaces, a certain amount of energy is absorbed by the free electrons in the material. When this energy reaches a high enough value, a *photoelectron* is produced. Today, photoelectric cells have been made quite sensitive to light variations; however, their use is quite limited in the missile field. The photoelectric cell is limited in use because the light-energy source is interrupted fairly easily, thus making the photocell inoperative. Development of photoelectric cells that extend into the infrared region alleviates the problem of light interference. However, this type of cell is not then a pure photoelectric device but trespasses the thermoelectric field. To insure proper operation of a pure photocell, there must be a high degree of uninterrupted contrast between the target and background at all times.

One of the most common types of light seekers is illustrated at the right. It is composed of two elements: a light-sensitive cathode and an anode. The elements are covered with a clear glass bulb having the physical size of a conventional radio tube.

As light impinges upon the surface of the cathode, the cathode emits photoelectrons. These electrons are collected by the positive anode, causing a change of current to exist within the circuit. The anodes in these phototubes are generally placed at the focal point of the photo-emissive coated cathode.

Tubes of this type are used to a great extent in motion picture projectors and seeing-eye devices.

This is a good place to discuss television systems since they may be thought of as light-sensing devices and thus should be mentioned in relation to light seekers. An advantage of television is that it can be



Photoelectric cell

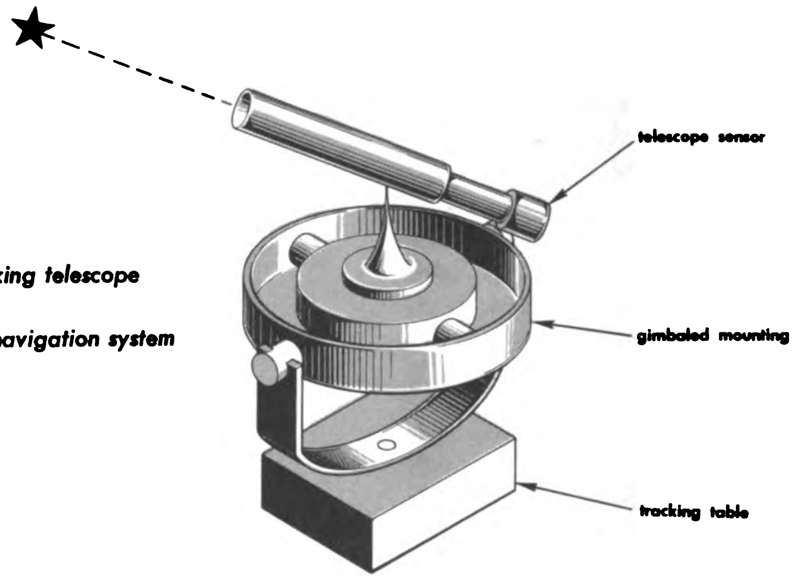
recovered to its course after being interrupted by some overcast. A serious objection to the television system is the difficulty of estimating the range of the missile from the target.

Television, as in the case of a photoelectric cell, is extremely susceptible to weather conditions. For accurate guidance, the picture should show instantaneous heading or ultimate destination, both of which are hard to attain because the angle of vision is limited and it is hard to compensate for wind and overcast variations. Accuracy is improved greatly by the use of an infrared sensitive device to enable TV to see through overcast.

Overall, the TV system has three principal disadvantages: (1) technically complicated, (2) susceptible to interference and jamming, and (3) the target must be optically visible.

The Iconoscope and Orthicon are two of the principal television pickup tubes used in the missile field. Tests are being continued to develop pickup tubes that will be of greater sensitivity, thus increasing the range and accuracy of the television system of guidance.

Of the more recent systems of guidance is the celestial navigation system. In this system, telescopes are used to sense the positions of selected celestial bodies. By determining the positions of bodies, spatial coordinates can be established. The coordinates are used as a guidance reference network. A set of telescopes which have a constant fix on at least two reference points is used.



Star-tracking telescope used in a celestial navigation system

This celestial system of guidance is considered a light-sensing device, the light source in this case being a reference rather than a target under consideration. The present systems of star-tracking telescopes are quite bulky; consequently, they are restricted to use in large missiles. This is a major disadvantage of the system, but it is used only for long-range missiles, which must be large anyway to accommodate the greater fuel supply needed.

Acoustic Sensors

The Navy has used the principles of sound detection for some time to determine the presence and position of submarines or other ships. Unless we consider a missile whose course is partly underwater at reduced speeds or at subsonic speeds in the air, this system must be ruled out. Naturally, no reliable acoustic signals are available at missile velocities greater than the speed of sound. Also, a system such as this would be greatly subject to jamming by the generation of interfering sounds.

Many problems arise when the use of an acoustic system of guidance is considered for air targets. The missile itself may generate so much noise that it would be impossible to follow any enemy aircraft. There is also generation of noise by the wind as the missile passes through the air. Filter systems, capable of picking out the target apart from the mother ship, are necessary to eliminate the possibility of the missile turning around like a boomerang and homing on its mother aircraft. One of the greatest limitations on such a guidance system is the background noise generated in the receiver itself due to the vibrations of the missile. To employ the acoustic system, as said before, it must be used on subsonic missile.

It is feasible to think of an acoustic system that would be able to intercept a missile that is traveling greater than the sonic speed. This, however, would involve a computing mechanism that would determine the speed and course of the craft and calculate a point of interception. The intercepting missile of this type would be quite complicated compared to other systems that have already been devised for supersonic missiles.

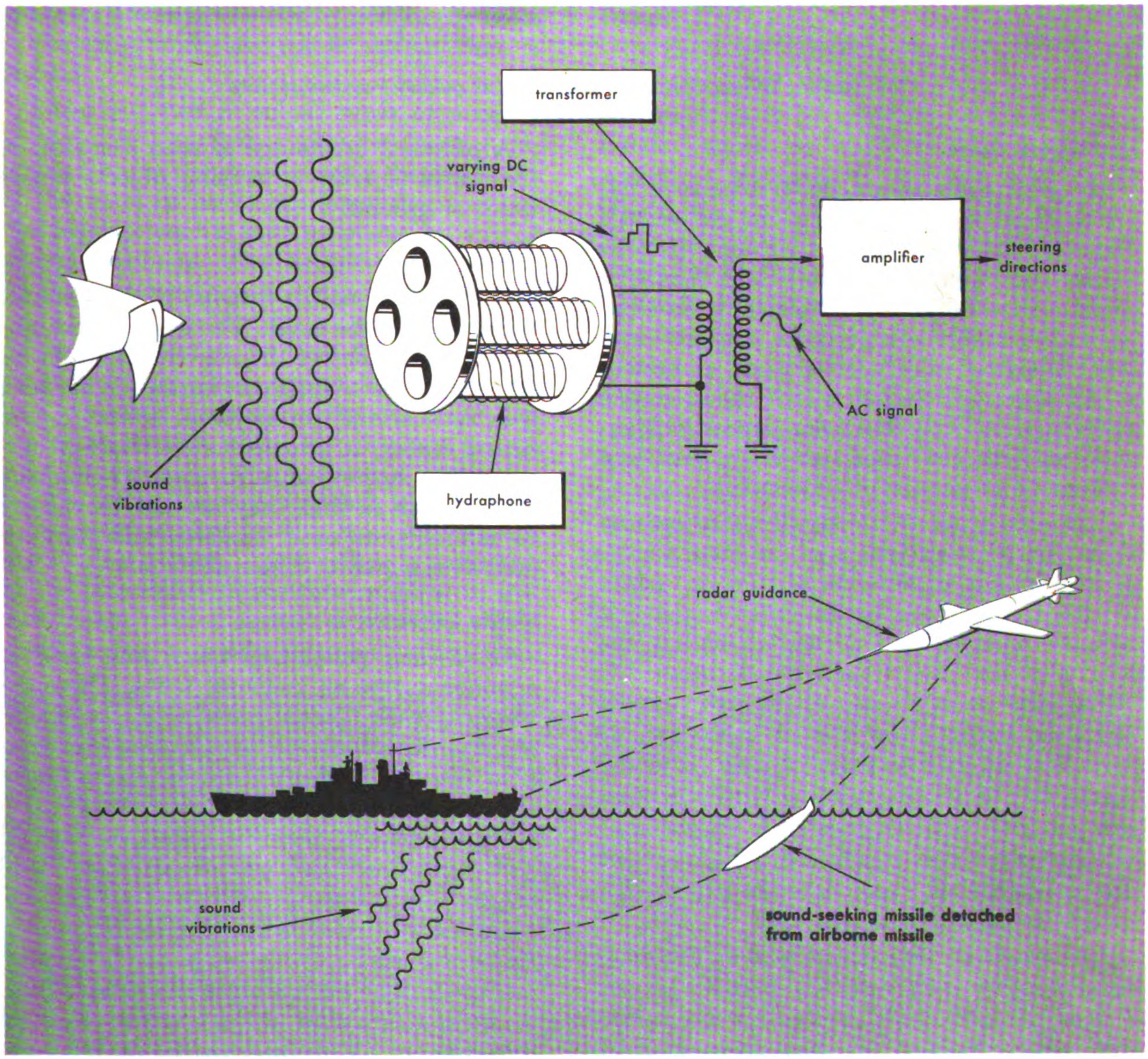
Sound-seeking devices have been used with success against underwater targets for the terminal phase of guidance. In these cases, a *hydrophone* was used as the sensor unit. A hydrophone system, shown at the right, senses the vibrations given off by an underwater target. Vibrations from the target set up alternating currents in the coils of the hydrophone. A sensing-element pickoff, generally a transformer, increases the amplitude of the signals and passes them on to further amplification stages. After sufficient amplification, these signals are passed on to an amplifying circuit that continuously looks for

a stronger signal. As the system looks for a stronger signal, it causes the missile to search out the source of sound energy.

Other devices used for collecting sound energy in water are microphones, magnetophones, and acoustical receivers. These sensing elements may be acoustical or electrical, for low or high frequency, resonant or nonresonant, rotatable or fixed, focusing or nonfocusing.

On the following page you will find a chart showing a comparison between the various types of sensors taken up thus far. It should be remembered, in referring to this chart, that no definite conclusions can be made as to what sensor unit should be used on a particular missile flight. The limiting factors of one system may be outweighed by its particular advantages.

Let's now consider some of the more recent types of sensing units. Some of these newer types may prove to be more useful on future missiles than the ones already discussed. Such will be the case, undoubtedly, for long-range missile development, since they are not limited by locally generated energy.



A sound-sensing system using a hydrophone

Type of SENSOR	Maximum Range	Information	Accuracy	Operating Conditions	Type of Target	Relative Speed of Target
Radio	Horizon, more or less depending on power and frequency	Direction	Good	No restrictions	Must carry transmitter	No restrictions
Radar	Horizon or less	Direction and range	Good to excellent	No restrictions	Few restrictions	No restrictions
Heat	Horizon or less	Direction	Moderate	No heavy fog	Special	Moderate
Light Television	Horizon or less	Direction and range	Moderate to poor	Clear daytime or flares	Simple contrast	Moderate
Sound	Short	Direction	Fair	Minimum background noise	Noisy	Small in comparison to speed of sound

Factors limiting usage of some types of guidance-system sensors

Magnetic Sensors

Magnetic systems are under extensive developmental tests. Variations in the earth's magnetic field, effects of magnetic disturbances, and difficulties in dip and magnetic components add to the problems of developing a dependable and accurate magnetic system. However, the use of such a system would lend itself well to a long-range missile and would be quite free from jamming. Magnetic devices are used often as auxiliary equipment to various other systems.

Magnetic systems employ flux-gate sensing units that sense variations of the earth's magnetic field due to deposits of iron in certain localities. Variations in the earth's magnetic field are sensed as flux density changes in the flux gate and are transferred into an AC signal that positions the missile to a predetermined course. The use of the magnetic system is limited to regions away from the poles. This limitation exists because the polar regions have frequent magnetic storms that would disrupt the system. Also, the magnetic field about the polar regions is not well enough defined for accurate guidance.

The most common sensor used in the magnetic guidance system is a *magnetometer*. Generally, three such devices are positioned mutually perpendicular to each other in order to sense the missile's position along a fixed magnetic line.

A principal advantage of such a system is that it is capable of sensing objects under water or ground.

Gravitation Sensors

Gravitation sensors respond to the gravitational field of the earth. The system does not use a single device but a combination of sensing devices that give an indication between true vertical and gravitational down. As was stated earlier, a principal advantage of systems that operate due to some phenomena of the earth is that they are not as subject to jamming as other systems and may be used over a greater distance.

Inertia Sensors

Gyros and pendulums have been used to a minor degree as guidance units. These devices, already highly developed for many other applications, are undergoing considerable research and development to improve their characteristics as guidance sensors. Gyros are generally associated with the control section of the missile and were taken up in detail at that time.

To detect the degree and direction of trajectory change, use is made of accelerometers. An accelerometer is an inertia device. A simple illustration of the principle involved in accelerometer operation is the human body in an automobile. If the automobile is subjected to sudden acceleration, the body is

forced back in the seat, and if the auto is stopped suddenly, the body is thrown forward. When the auto goes into turns, the inertia force on the body is away from the turn.

The field of accelerometry is not new, and a great number of devices for measuring static and dynamic acceleration are in use. The basic principle of operation of an accelerometer consists of the measurement of the inertial reaction force of a mass to an acceleration. Many different methods can be used in measuring this force. The choice of the method is mainly dependent on the frequencies present, whether or not acceleration is changing, and the type of output that is desired.

There are two principal types of accelerometers. In the first type, the inertial reaction force of the mass causes a displacement of the mass in an elastic mounting system. The displacement is then measured by any of several methods. The second type of accelerometer operates on a fundamentally different principle. The force that counteracts the inertial reaction force of the mass is not supplied by an elastic mount but is supplied by an electric current, a stream of air, or any other system which can produce a controllable force. In this system, a small deflection of the mass is detected, and a force is instantly applied to the mass to prevent any further motion. Acceleration is indicated by the magnitude of the force applied to the mass by the electric current or whatever other means is used to produce the balancing force.

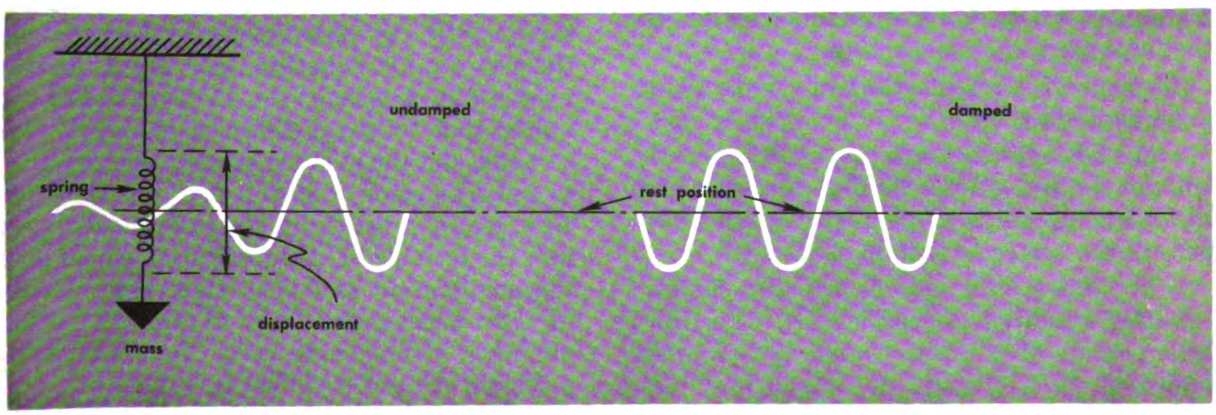
The following explanation is given in order to help you to better understand the fundamental principle of accelerometers:

Whenever a mass which is supported by an elastic mounting system is displaced due to some force, it tends to oscillate about its rest point. This oscillation is known as *simple harmonic motion*. The number of oscillations during a period of time are limited by *damping* action of the system. Damping limits the amplitude of each oscillation; thus the number of oscillations is limited. The damping action may be increased by the use of a device especially for this purpose, called a *damper*.

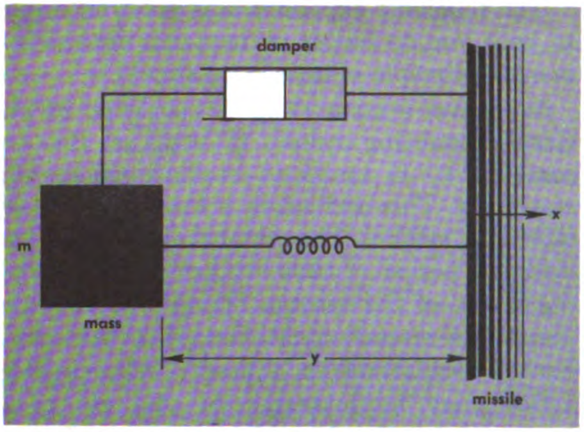
In the figure below, if the spring-suspended mass were displaced and then released, it would tend to oscillate in simple harmonic motion. If there were no retarding action by the spring, it would oscillate as illustrated in the "undamped" condition. However, a spring does offer resistance to the oscillation; consequently, the oscillations will diminish as represented in the "damped" condition.

In order to have an efficient accelerometer, it is necessary that the amount of damping applied to the system be just great enough to give a nonrestricted displacement and yet prevent any oscillations from existing.

Consider the accelerometer on the next page. If the missile suddenly accelerates some distance in terms of "X," the mass "m" is displaced (relative to the missile) some distance in terms of "y." The relative accelera-



Effect of damping on a spring-suspended mass



Accelerometer's function in inertia-sensor unit

tion of the mass during this time indicates a force present because:

$$F = Ma$$

where "F" equals force, "M" equals mass, and "a" equals acceleration.

There is a certain displacement of the spring per unit of force applied. The spring at the end of this period of time tends to reverse the direction of force on the mass. This would lead to simple harmonic motion if allowed to stand alone. Therefore, the damper is used to produce an action that limits this oscillation. We can say that the damper must attain a certain velocity at the end of this time so that its retarding action will counteract the force set up by the spring as a result of the displacement of the mass. The damper will again present such a counteracting force when the mass has reached its maximum displacement in the opposite direction. Displacement in the opposite direction is due to the force exerted by the spring.

There are three results that may come about in an accelerometer:

a. The system may be *overdamped* (it will not oscillate). This condition occurs when the damping action is greater than the tendency of the spring to sustain oscillation.

b. The system may be *critically damped* (limiting condition of oscillation). This condition occurs when the above-mentioned forces are equal.

c. The system may be *underdamped* (will oscillate with decreasing amplitude). This condition occurs when the damping action is less than the tendency of the spring to sustain oscillation.

Under certain conditions, the displacement of the mass relative to the missile can be used to measure one of three things:

1. The displacement proportional to acceleration.

2. Displacement due to the velocity, as indicated by the damping action.

3. Displacement proportional to distance moved by the missile.

For our use, we wish to measure acceleration; we want the accelerometer to operate in the underdamped region, bordering the critically damped condition. This gives us a system that is sensitive, and at the same time the system tends to prevent any transient oscillations. The accuracy of the system depends on the method used to measure the displacement of the mass relative to the missile and on the linearity of the elastic mounting. The sensitivity of the system may be increased by reducing the frictional forces in the mounting and by improvement of the indicators used.

As you learned in the chapter on physical principles, the average acceleration required to displace an object a distance "s" in "t" seconds, assuming zero initial velocity, was given as:

$$s = \frac{1}{2} at^2$$

where "s" is the distance in feet, "a" the acceleration in feet per second per second, and "t" the time in seconds.

The accelerometer sensitivity required for accuracy in a given system may be found by the use of this equation. For example, if a missile has a velocity of 1500 mph, the required time to travel 400 miles is roughly 960 seconds, and the average acceleration required to cause an error of 3000 feet in this time can be found by substituting the time and distance into the formula. This substitution would give the acceleration as 0.0065 ft/sec/sec. Therefore, the acceleration being in error by 0.0065, all other errors being zero, there would be a system error of 3000 feet in 400 miles.

As you can see from the example, the acceleration taken here was quite low. Ac-

celerations as great as 500 feet/sec/sec may have to be measured. Therefore, the accelerometer must be capable of satisfying either extreme of conditions. This means that the accelerometer must be capable of measuring one part in 100,000. At present, accelerometers capable of measuring one part in 1000 are available. Much work and research still need to be done in order to give the sensitivities required of accelerometers.

Shown on the right is the second mentioned case in which the force that counteracts the inertial reaction force of a mass is not supplied by an elastic mount but is supplied by an electrical current.

Let "e" be a voltage that is developed due to the displacement of the mass relative to the missile and "i" the output of the amplifier (the restoring current fed to the coil).

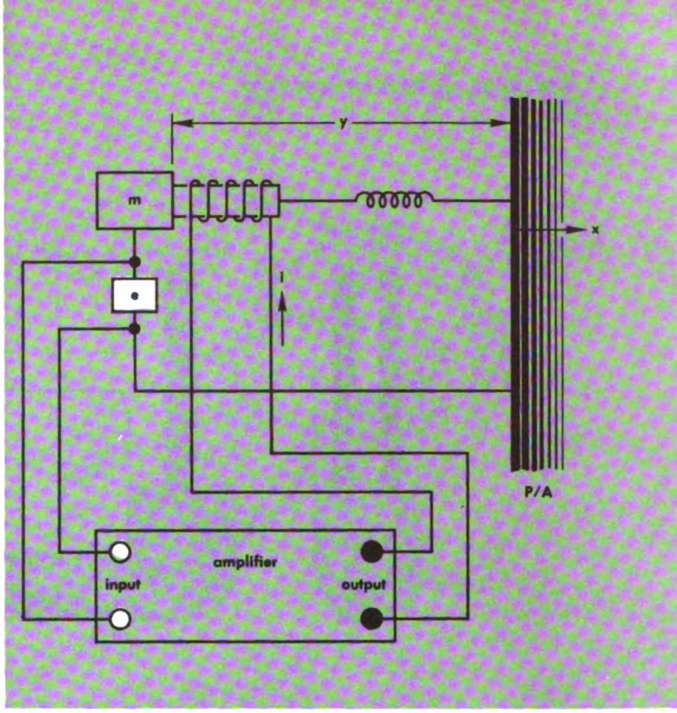
If "e" is dependent on the displacement of "y" of the accelerometer mass (m), then the output of the amplifier (i) is dependent on the displacement (y). If the missile accelerates, there is a certain voltage applied to the amplifier input. After necessary amplification, there is a proportional output applied to the coil in such a manner as to counteract a certain amount of the displacement. As the mass (after initial acceleration) tends to move in the opposite direction, another restoring current is generated.

The acceleration, therefore, could be measured by the displacement, by the input to the amplifier, or by the restoring force of the current fed to the coil.

The displacement allowed to the mass is small, which means that errors which may arise due to the nonlinearity of the electric field are mostly eliminated. The sensitivity of such a system can be increased with little effect on the range of operation.

There are many ways of obtaining a signal. Obtaining a signal is a function of the displacement of the mass in the first-mentioned type of accelerometer.

Another method has been developed by the Massachusetts Institute of Technology. This method utilizes a change of inductance between two coils as their separation is varied. The coils form part of an inductance bridge. As the inductance changes, the bridge is unbalanced and an AC voltage output results.

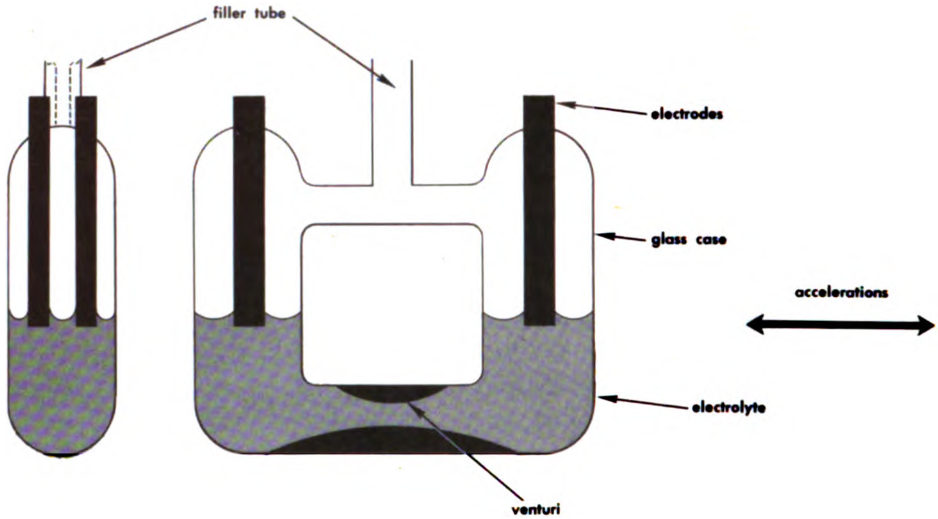


Accelerometer in which inertia reaction force of a mass is supplied by an electric current

Another accelerometer uses wire strain gauges as the suspension elements for the mass. The strain gauges form the arms of a bridge. A change of acceleration causes a change of the electrical resistance of the circuit, giving an AC output that is an indication of the acceleration.

An interesting example of the second type of accelerometers is the vibrating accelerometer, which was developed at Fort Bliss. In this accelerometer, the mass and coil vibrates, causing periodic currents to flow through a coil. An acceleration causes the average momentum of the system to change. The average current output of the system is proportional to the average acceleration.

Still another principal type of accelerometer is the manometer. In this type, accelerations are controlled by the electrolyte flowing toward one or the other ends of the manometer. This action provides current control between pairs of electrodes. The venturi, shown in the following figure of a manometer, damps the response of the manometer by controlling the electrolyte movement. This damping prevents oscillation of the electrolyte, which in turn prevents oscillation of currents fed to the control system.



Manometer accelerometer

A photo-optical type of accelerometer is under development. It has the disadvantage of being relatively large for the purpose. It requires a partially stabilized light source, a reflector and lens system, a light valve, and a phototube, all of which have to be gyro-stabilized since they form a part of the converter.

Another accelerometer in use is of the vibrating-reed type. In this type, a change of frequency indicates the acceleration along the axis of the reed. The reeds are used, facing in opposite directions, so that the difference in frequency between the reeds is proportional to acceleration in either direction. Tuning forks have been used instead of simple reeds to eliminate dead spots when the reeds are vibrating at the same frequency. The two tines of a tuning fork always vibrate 180 degrees out of phase, and the energy of vibration, therefore, would be theoretically cancelled out at its base.

Accelerometers in which the capacitance of a circuit is varied proportionally to the acceleration also are employed.

There are many other ways of measuring the acceleration of a missile. The methods just discussed are just a sample of the many ways used to accomplish this function.

It should be remembered that when a guidance system is chosen for a missile, various advantages and disadvantages of each system must be considered before making a selection. Also, it may be necessary to use a combination of two or more of the systems to attain guidance throughout the flight. No matter what system or systems are chosen, the sensor element is the *eye* to an otherwise blind flight.

SIGNAL PICKOFFS FOR SENSING ELEMENTS

In order to transfer an incoming signal from the sensor unit to later stages of amplifying and computing, use is made of a sensing element pickoff. The sensing-element pickoff, as you learned in the preceding chapter, is any unit or combinations of units that take a received signal and transfer it in a usable form to the following stages.

Sensing Element Pickoff for Transferring Regular RF Frequencies

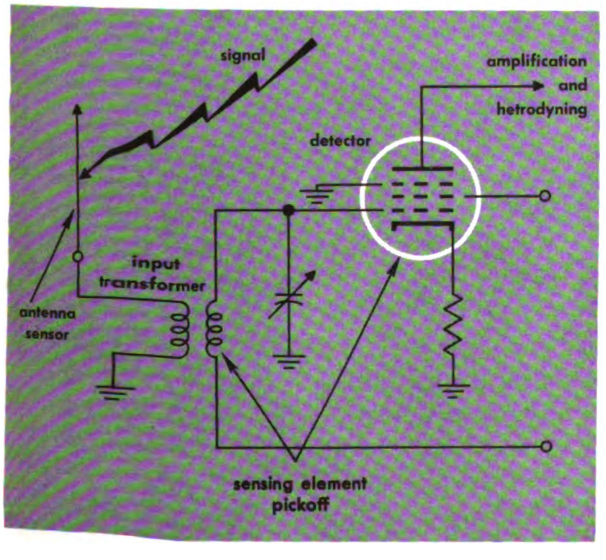
In the case of transferring regular RF frequencies, the usual type of sensing-element pickoff is the transformer. The primary of the transformer is connected to the sensing element, and the load is connected to the secondary. An *input transformer* refers to a transformer that is used to couple a low-impedance source of energy to the grid of a tube. The input transformer generally has a high step-up ratio so that the magnitude of the incoming signal will be such that it may be amplified in later stages.

In the RF system of guidance, an antenna is considered as the sensor, and the unit that relays the signals to later stages is the sensing-element pickoff.

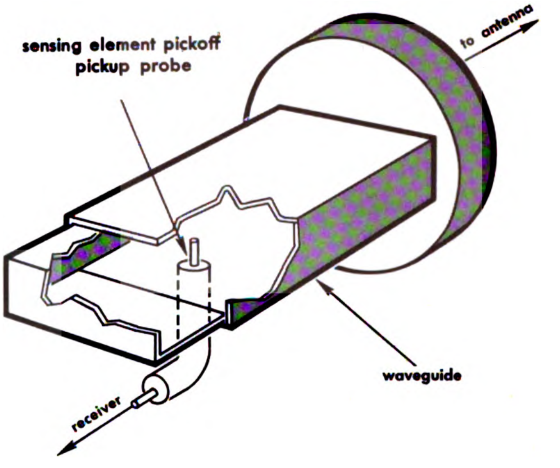
In considering the figure below, note the dotted line which indicates the *detector* as a portion of the sensing-element pickoff. It is often difficult in an electronics system to clearly define the limits of a particular unit. You may find it easier to visualize the detector as the pickoff unit rather than the input transformer. Or perhaps we should consider both the input transformer and the detector as the sensing-element pickoff. No matter what you choose to call them, the important point is that the sensing-element pickoff takes from the sensor a received signal and transfers it in a usable form to the following circuitry of the system.

Pickoffs Used in Radar Systems

In a radar system, the energy that is received by a waveguide after leaving the antenna is transported along the waveguide to a point where it must be transferred once more to a usable form by the receiver. This energy may be fed directly into a resonant cavity where oscillations are initiated to excite amplifying stages. The energy also may be detected by crystals or tuned oscillator stages.



Sensing element pickoff used for transferring regular RF frequencies



Waveguide section and pickup probe

If the energy is to be collected from the waveguide, use is commonly made of a pickup probe. Shown above is a cutaway sketch of a waveguide section and pickup probe. A probe will receive a maximum of energy if it is inserted into the waveguide at a certain point and position.

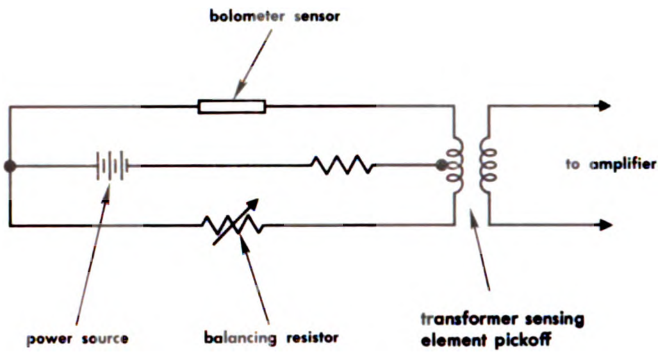
Thus, the energy that was returned from the target is available for the receiver. The receiver accomplishes the heterodyning, detecting, amplifying, and computing necessary to cause the target selector to operate at the proper time for providing a signal to the controls unit.

The amount of energy available to the receiver is proportional to the effectiveness of the antenna sensing unit.

Let's note once more that clear definite limits are difficult to determine when referring to the sensing-element pickoff. In the present case the waveguide, the pickup probe, and possibly a tuned circuit may be combined as a unit and referred to as the sensing-element pickoff.

Pickoffs for Infrared Sensing Elements

The infrared types of sensing elements have a number of different types of pickoffs as is true in the case of RF and radar systems. Common use is made of a bridge circuit as part of the sensing element pickoff. The bridge circuit is useful in the heat-detecting systems in that sensors for these systems



Typical bridge circuit used with a bolometer

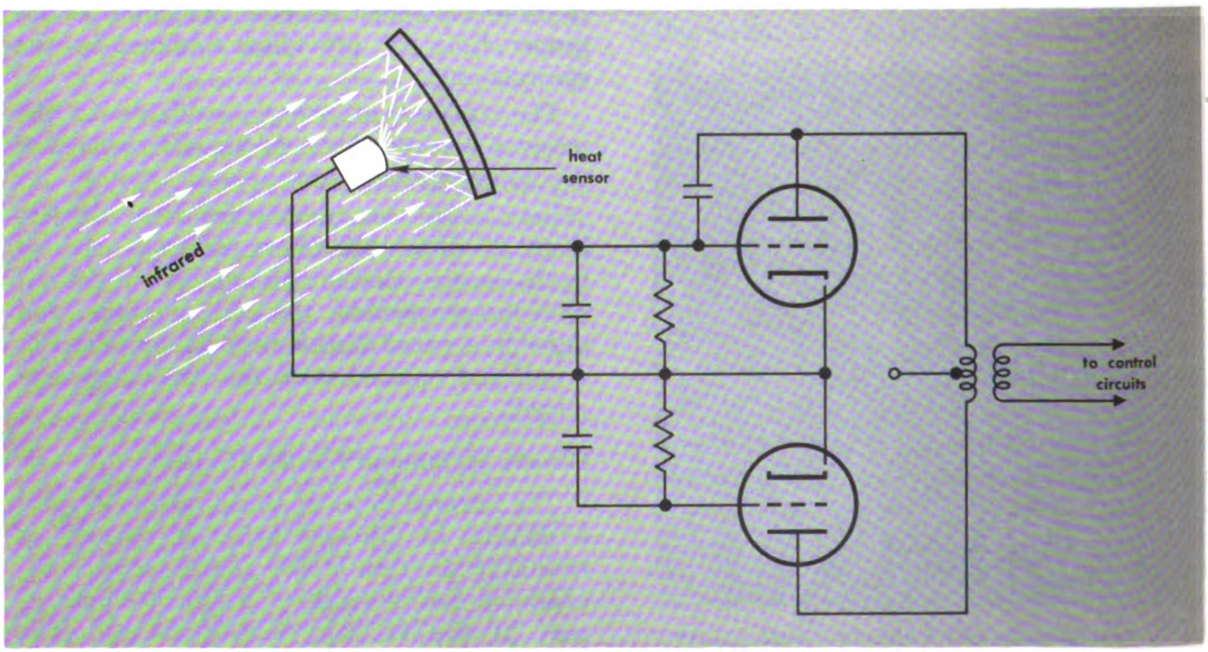
generally operate on the principle of resistance changes.

Shown above is a typical bridge circuit used with a bolometer sensor. The changing temperatures that are applied to the bolometer are accompanied by a fluctuating resistance so that the bridge circuit is thrown off balance. The transformer forming two arms of the bridge acts as an impedance match between the bolometer and the high impedance of the input to the amplifier. The transformer also converts the signal to an approximate sinusoidal waveform before passing it on to further amplification stages.

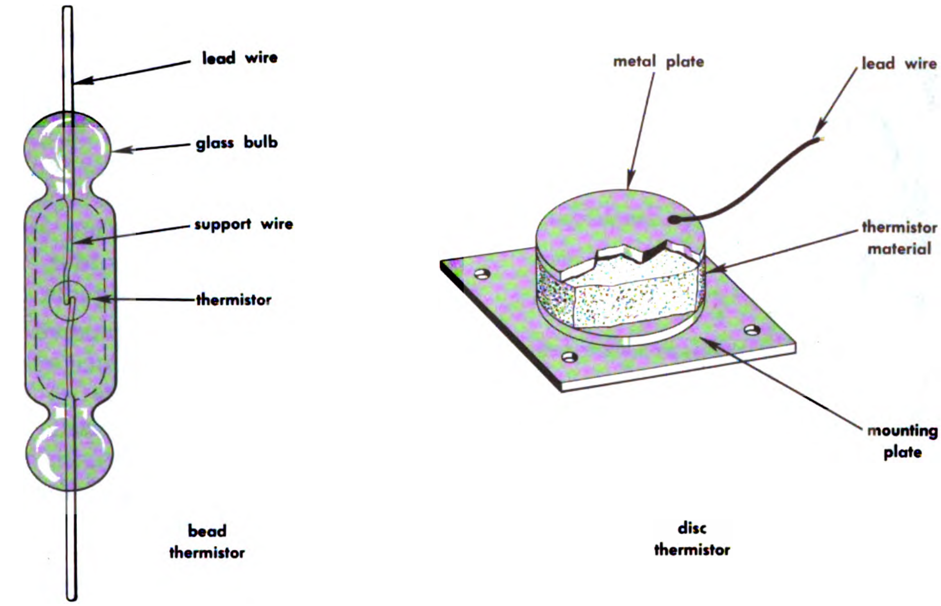
In this case the fundamental oscillator frequency is caused to change, giving an indication of the magnitude of the received infrared energy.

As in any of the sensing-element pickoffs, the infrared pickoff should be sensitive enough to quickly relay any received change of signal and be subject to a minimum of operational difficulties.

Some types of sensing-element pickoffs were mentioned previously in the chapter on control system components (Chapter 5, Section 1). It may be found that some of these could be applicable to guidance systems as



Oscillator sensor pickoff



Thermistors

well. One such device, for example, is the *thermistor*.

A thermistor is a variable resistance element. The resistance of a thermistor varies inversely as the temperature; that is, as the temperature increases, the resistance decreases. The resistance of the thermistor is varied by thermal changes in the environment, by AC or DC flowing through it, or by absorbed RF power.

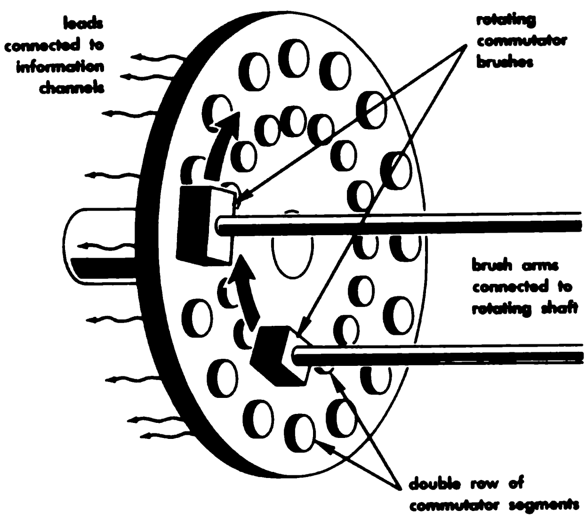
Two types of thermistors are illustrated above. The *bead thermistor* used for power measurement has a small mass and is affected by the changes mentioned in the previous paragraph. The *disk thermistor*, used in compensating networks, has a large mass, and its resistance is relatively unaffected by the current flowing through it. Its resistance is dependent primarily on the ambient temperature.

In the case of these thermistors, the bead thermistor could possibly be used in RF systems as a sensing-element pickoff unit as well as in a control system. The disk thermistor would find its primary use in control systems.

This common use of units in both the guidance and control systems brings to light another situation that is common when dealing with missile systems; that is, the line of demarcation between the control function and the guidance function is, quite frequently, narrowly and arbitrarily drawn. The two functions are always dependent on each other and closely integrated, and the problems that arise in one may not be analyzed properly without due consideration to the other. It cannot be overemphasized that guidance and control must function as a unit in order to have a satisfactory missile flight.

Commutation Principles

In conjunction with many of the sensing-element pickoffs, there is a device which connects the information channels in the circuit to the receiver at periodic intervals for relatively slow cycle information. In any guidance system there are numerous information signals as to range and altitude being received, and it would be impossible for the guidance system to distinguish between the various received signals without first segregating them before passing them on to the remaining amplifying stages. Since the order in which the channels are sampled is known, a *commutator* can be used to separate the information so that it may be interpreted by the receiver.



Operation of commutator

The commutator in this case, contrary to your knowledge of motors and generators, is the stationary member while the brushes are the rotating part. Information is fed into the current-carrying segments of the commutator as shown in the figure above and then picked-off by the rotating brushes. This gives a means of systematically measuring and transmitting the measured information so that various factors may be indirectly observed by a ground-control operator.

A synchronizing pulse usually is used to indicate to the receiver the starting and stopping point of a cycle. Commutators are usually driven by a geared-down dynamotor in order to have a smooth operating system. By gearing the commutator, voltage changes felt by the dynamotor are not felt as seriously by the commutator. For example, if a dynamotor is operating at 9000 rpm and the commutator is geared for 300 rpm operation, the voltage changes appearing at the dynamotor are proportionally reduced at the commutator.

Keep in mind that the pickoffs discussed in this chapter are just representative. There are other types of pickoffs, and some of the ones mentioned here with respect to a particular guidance system may be used on other guidance systems. The major consideration in choosing a sensing element pickoff is whether or not it is electrically applicable to the system.

So far, we have discussed the means of sensing an incoming signal and relaying the signal to other circuitry. These units that have been discussed generally are not limited to one-way operation; that is, they are not used solely for receiving an energy signal. They also are used to transmit a signal. The basic units used for radiating energy would have similar devices in order to give off information or guidance signals.

In order to give off or receive a signal by one of the particular guidance systems, there must be a systematic way of determining the relative position of the missile to the target and/or the base control station. The way in which the sensor "looks" for its position is referred to as *scanning*.

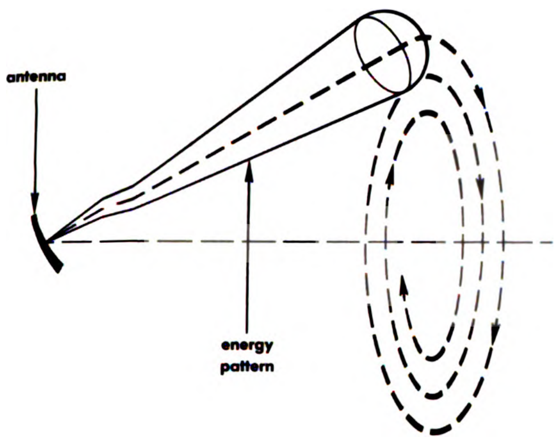
SCANNING METHODS

Effective search for a target or a missile that is to be guided can be made only when the whole area of interest is canned without gaps. Therefore, the search must be rapidly carried out in some methodical manner. A radar system may have a line-of-sight path and a maximum range far exceeding the range of a missile and yet be useless if it fails to pick up, track, and guide the missile early in its trajectory.

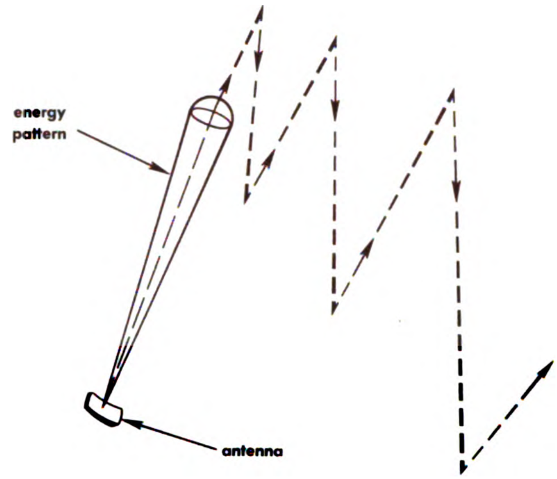
The scanning procedures discussed here are not limited to radar systems alone. Scanning procedures may also be employed in infrared, light, or any other system that focuses a beam of energy.

For limited solid-angle search, as would be the case when a missile is "lost" due to intermittent jamming or fading, a *spiral* or *saw-tooth* scan can be used to cover the volume of space. A beam which does not fill either dimension of the solid angle covers the volume of space.

A common type of scanning used in beam-rider systems is that of *conical scanning*, shown on the right. Conical scanning occurs when a beam is slightly off center and rotates in a manner which generates an elongated football-like volume. The signals are strongest within the generated volume. The longitudinal axis of the beam passes through the radar antenna and the target. When a missile travels along the axis of the "football," all



Spiral scanning



Sawtooth scanning

signals from the rotating beam are neutralized in the detector carried within the missile. However, should the missile get off the axis of the beam, the signals on one side would become stronger in proportion to the other side. This condition activates a mechanism to steer the missile back on the axis.

In order to obtain 360 degrees of azimuth search, use is made of the helical-type scan shown on the right below.

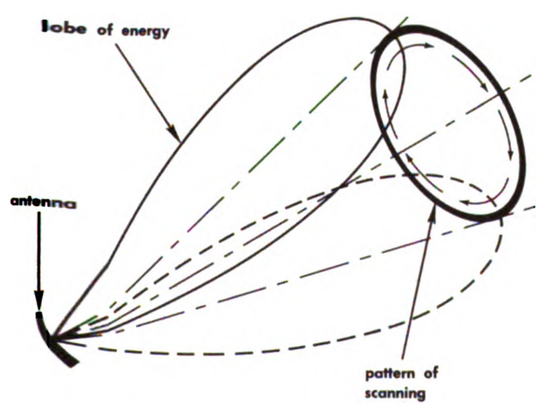
In the helical scan, the antenna tilts one-half beamwidth in elevation for each complete revolution in azimuth. Normally, the scan is not carried above 30 degrees in the vertical direction because a large portion of the search time would be spent in this region. This region represents a small portion of the volume of space in which objects may be found.

In order to insure solid coverage, the separation between turns of the scanning cycle should be approximately half of the beamwidth. This means that each succeeding scanning cycle will overlap the previous one. In the case of the spiral scan, the radial motion must have a fixed relationship to the circular rate; and the horizontal motion of the sawtooth must be related properly to the vertical scan speed. The main factors to consider in a scanning process are:

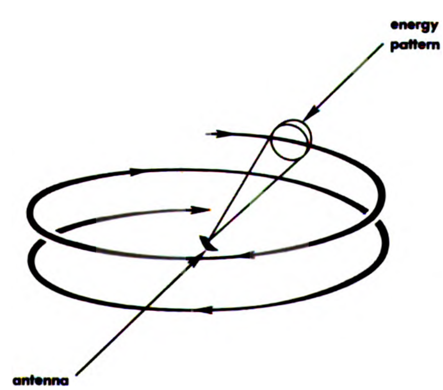
1. The minimum scan period required for a given solid angle of search at the maximum range.

2. The maximum angular velocity at whcih the beam will successfully search at the maximum range.

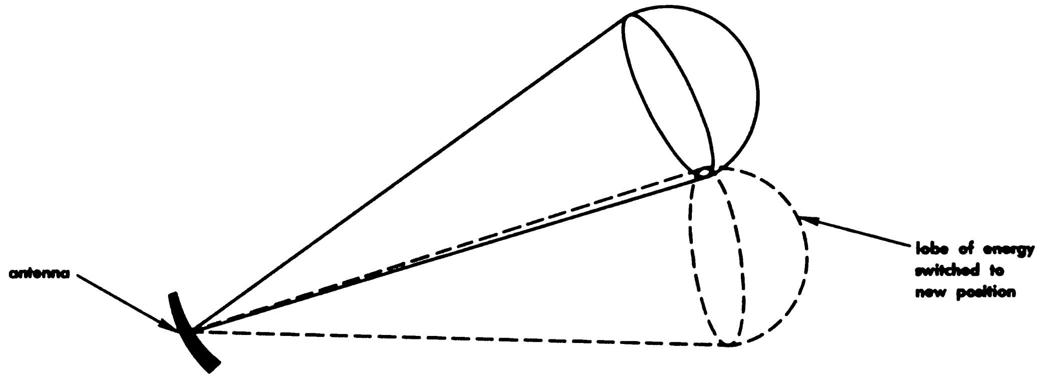
3. The distance that a missile flying away at a certain speed along the line of sight of



Conical scanning



Helical scanning

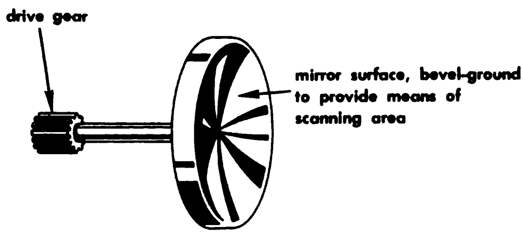


Lobe switching

the radar will travel between two successive scans during search at maximum range.

Another means of positioning a missile with respect to a transmitter is by lobe switching. In this system, two overlapping beams are used to increase the angular deviation of the beam. Note the illustration below. The lobing is accomplished by using two different antenna systems or by the use of one antenna system which is made to oscillate by mechanical or electrical means.

Various devices are used to obtain the desired type of scan for the guidance system. In the radar systems the lobing is accomplished by shaping the reflector and then causing it to rotate about the fixed antenna. A similar pattern can be attained by keeping the reflector stationary and causing the radiator to rotate. Offset antenna patterns can be attained also by the lobe-switching method.



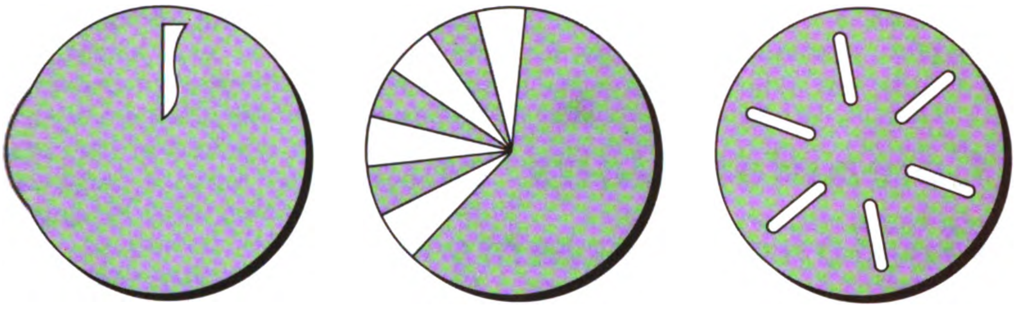
Vibrating mirror used for scanning

For the infrared and light-seeking devices, use is made of vibrating mirrors in order to scan the area. As a mirror rotates or vibrates, it causes the focal point to shift about the sensing element, thus giving an indication of the point where the target signals are the greatest.

For star-tracking systems and some of the light-seeking systems, split or slotted disks are used in order to interfere with the light at a definite rate. The use of a split disk fixes a definite area to be scanned by the system and permits time for the circuit to recover from radiations that were scanned on the previous rotation. Pictured on the right are various types of slotted disks used in missile systems.

POLARIZATION METHODS IN MISSILE GUIDANCE SYSTEMS

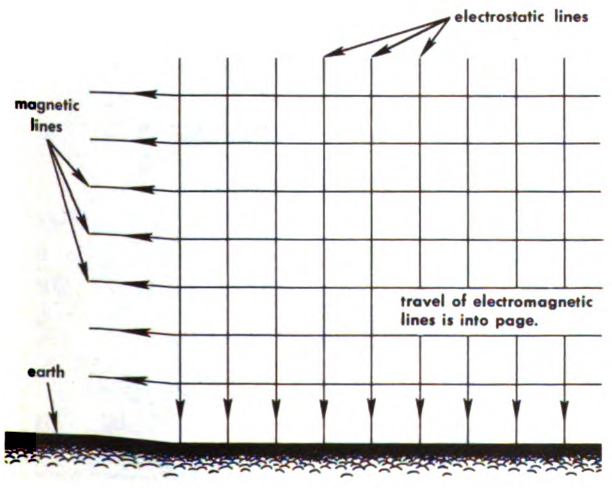
Polarization of antennas is important in missile guidance systems. As stated earlier, polarization has to do with the direction of the electric field. Electromagnetic waves are made up of two fields: the electrostatic and magnetic fields. The electrostatic and magnetic lines of force travel at right angles to each other in a plane that is moving in a direction perpendicular to the lines of force. The arrows in the figure on the right indicate an instantaneous direction of the fields of a vertically polarized wave. The direction of travel of the fields is into the page. If either of the fields were to be reversed, the direction of travel would be reversed.



Slotted discs used in star-tracking and light-seeking systems

The electromagnetic wave is *vertically polarized* because the electrostatic lines of force are perpendicular to the earth. The wave would be *horizontally polarized* if these lines of force were parallel to the earth. These two general types of polarization are referred to as linear polarization.

The electrostatic field of any wave tends to change polarization. This change is small at the lower frequencies but tends to be more pronounced as the frequency is increased. Since it does tend to change its polarization, it takes on a rotational characteristic of either sense. Due to phasing of portions of the waves when the field encounters a target, the sense of rotation may be in both directions. This is particularly the case when it encounters an asymmetrical target. The phasing in this case would tend to give an elongated field of electrostatic energy. The wave would be *elliptically polarized.*



Instantaneous direction of the electrostatic and magnetic fields of a vertically polorized wave

The field becomes *circularly polarized* when the phase difference of the electrostatic fields is 90 degrees. The field of electrostatic energy is circular in nature and may contain either directional senses.

When the polarization of a field is changing, it is difficult to detect a maximum of usable signal. However, if a device is used in which reception of either directional sense can be detected, the signal *definition* may be greatly improved. The general case in radar systems is that the symmetrical areas of scan will return a signal circularly polarized, while a asymmetrical area will return an elliptically polarized signal. The picture seen on a radarscope is greatly improved if both senses of field rotation are detected.

Much research is being done in the field of antenna polarization. Development of predictable waveguide sections (rectangular and circular) and phasing devices are under continuous study.

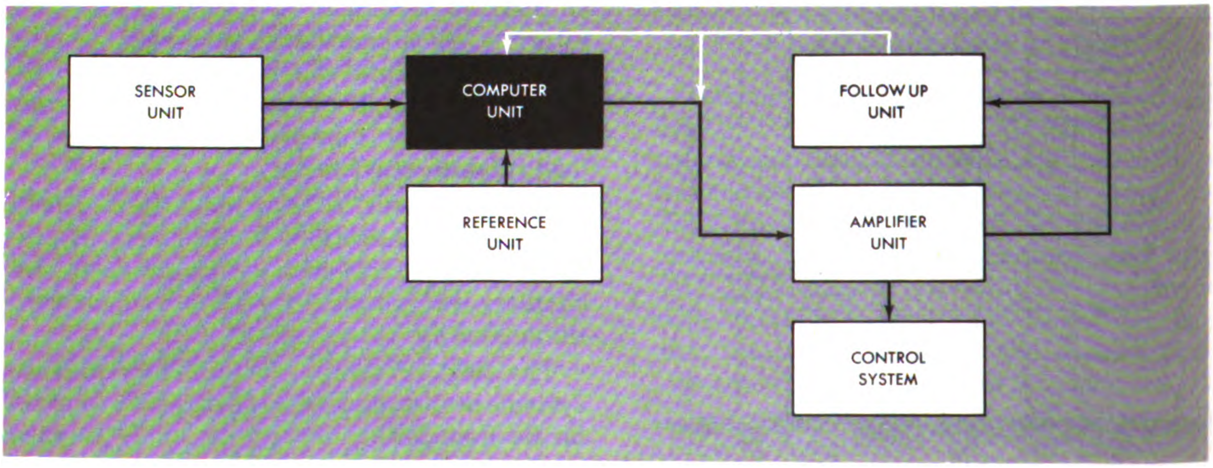
The field of sensing devices for guidance systems is vast, and we have here just touched on its boundaries. We're now ready to take up computer units, included in the next block of the guidance system block diagram at the beginning of this chapter.

Computer Units of Guidance Systems

Computer units in missile guidance systems are found in various forms. A computer unit may be simple, such as a mixing circuit in the airborne vehicle; or it may be complex, as in a large-scale ground installation where the complete flight of the missile is determined. Whether airborne or ground-based, the computer unit is related to the other guidance units as shown in the accompanying block diagram.

An important function of the computer unit in many guidance systems is the coding and decoding of information relating to the missile flight. Enemy countermeasures or the guidance of more than one missile at a time may make this operation necessary. *Discriminating* circuits are used to select pulses of particular width, amplitude, frequency, phase, and/or time difference, and to reject all others. Interference in the form of noise is reduced to a minimum by these decoding circuits in order to develop reliable flight signals.

Another function of computer units is the *mixing* of signals from sensor and reference units to produce error voltages. These signals are sometimes mixed in a certain ratio or according to programmed instructions. Error voltages are fed to the amplifier unit which increases their magnitude. Then the error voltages are introduced into the control system to correct any deviation from the computed flight path. The amplified error voltages also are passed through the followup unit to be reprocessed in the computer.



Basic missile guidance block diagram

There are components in some guidance computer units that compare two or more voltages in order to produce error signals. These components may be in the form of *voltage* or *phase comparator* circuits. The autosyn units that are discussed in the preceding chapter are used in guidance computer units to resolve signal voltages into components for processing. In some guidance systems, it is necessary to convert earth-reference coordinates to space-reference coordinates by autosyn resolver circuits.

Airborne computers may be classified in terms of phases of the missile flight. They may be in the form of distinct units or combinations of the following computer units:

1. Prelaunch computer.
2. Launch computer.
3. Azimuth computer.
4. Elevation computer.
5. Program computer.
6. Dive-angle computer.

Another method of classifying computer units is to consider the operating principle. Both airborne and ground-based units use two classes: *analog* and *digital*. Their major differences are presented briefly in this introduction. The tow operating principles involved in their individual component operation are discussed later in the chapter.

An analog computer manipulates physical quantities that represent the mathematical variables of the particular problem under study. In the mechanical type of analog computer, for example, the machine variables may be rotating shafts driven by gear trains. The angular displacement of the shafts is measured to produce the solution to the equation or mathematical operation. Computers for solving navigation and bombing problems have used these principles for years. Early radar gun directors also made use of this type of operation.

Differential analyzers in both the mechanical and electronic form have been applied to solve problems in the missile field. The first large-scale differential analyzer was operated at M.I.T. in 1930. This original machine used mechanical principles. Current types of *electronic differential analyzers* are used in almost every phase of missile dynamics to obtain information without constructing actual prototypes of the vehicle. Although this machine is often called an analog computer or *simulator*, the computing elements are not usually direct analogs of the physical quantities in the problem. They are analogs of the mathematical equations describing the problem.

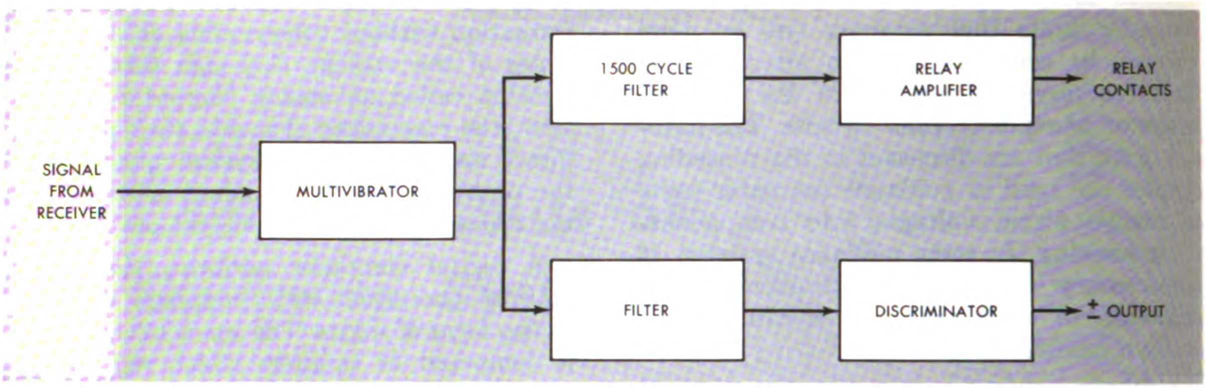
A digital computer performs the mathematical operations with numbers expressed in the form of digits. The machine essentially is composed of counters which register and add in discrete steps. Desk calculating machines are familiar forms of the mechanical type of digital computer. The first *automatic sequence-controlled calculator* used mechanical counters controlled by magnetic clutches and relays. The International Business Machine Corporation began the construction of this machine in 1939.

Electronic digital computers had their start in the *electronic numerical integrator and calculator*. This machine was the first to use electronic circuits as the actual computing elements. It was designed primarily for step-by-step numerical integration of the equations of external ballistics. Although this computer used a scale of ten, the airborne digital computers used in missiles operate on the binary system of numbers. The binary system is discussed later.

CODING AND DECODING UNITS OF MISSILE GUIDANCE SYSTEMS

Familiar examples of coding and decoding systems include teletype, transoceanic telephony, television, identification of friend or foe (IFF), radio, and radar beacons. In all these systems it is necessary to develop a synchronizing key or code at the receiving end that will properly match the coded transmission. This matching is necessary in order to extract the intelligence.

Types of coded transmission were described in chapter 4, section IV. Each type of pulse modulation — amplitude, width, rate, position, frequency, and count — requires a decoding circuit that reproduces the original signal and keeps interference and noise at a minimum. The decoding circuits described



Decoder for PRR system

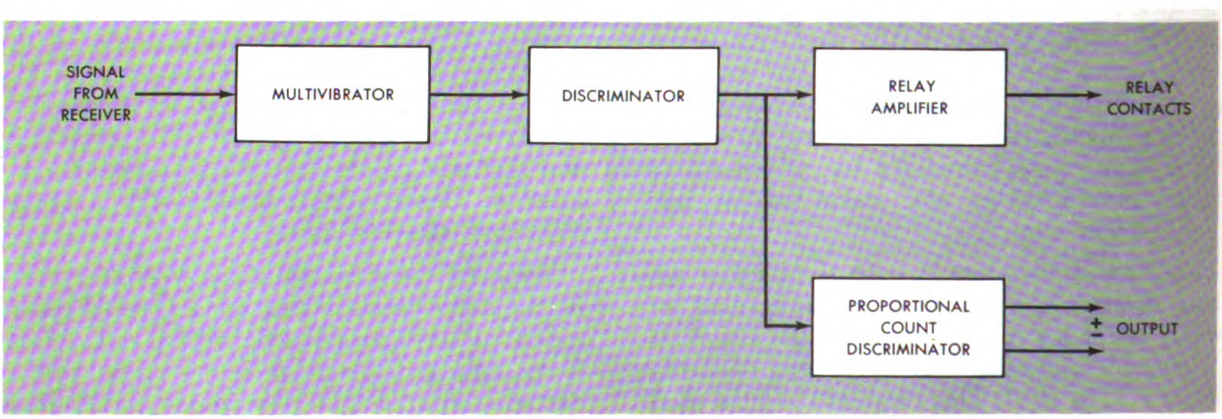
here are examples of those now being used in the missile guidance computer units.

The three systems described in the following paragraphs are used for decoding proportional information for lateral control of the missile and ON-OFF control for dump signals. Information is transmitted by discrete or continuous changes in pulse repetition rate, frequency modulation of the pulse repetition rate, and pulse time modulation.

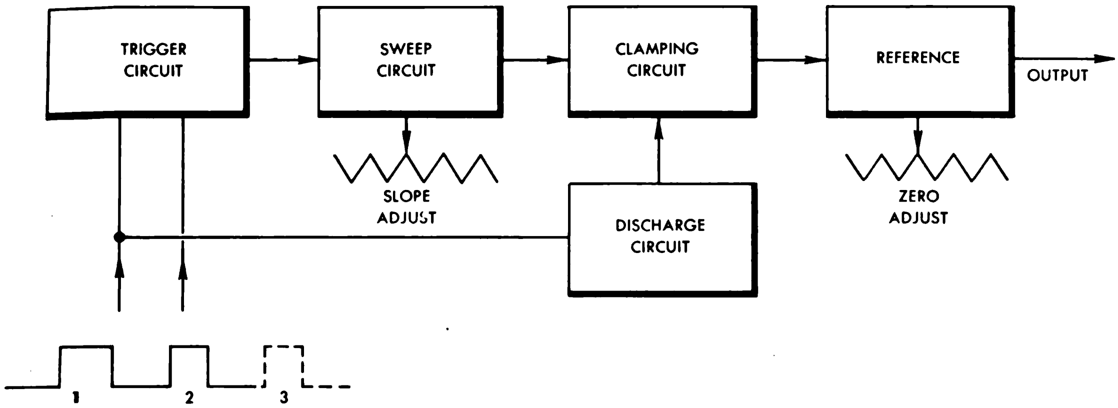
In the pulse-repetition rate system (PRR) which is diagrammed above, the coder determines the repetition rate of the transmission of radar pulses. For *proportional control*, the repetition rate varies from 1800 to 2000 pulses per second according to the DC error-voltage input. The on-off channel is operated by a preselected discrete pulse-repetition frequency (PRF) of 1500 cycles per second.

In a system using frequency modulation of the pulse-repetition rate, earlier referred to as pulse-frequency modulation (PFM), a subcarrier frequency modulates the radar repetition rate. Conventional frequency-modulation circuits are used for detecting the command information. Frequency instability is a problem in this system.

In the pulse-time modulation (PTM) system, three pulses are transmitted by radar. The first two are coded for beacon response. The third pulse is shifted in time according to intelligence being transmitted. At the receiver, the first two pulses are decoded and used as a reference and for beacon triggering. A radar is required that will supply three pulses, with the third pulse variable in time to provide the information for proportional control. This system is used in one of the latest command guidance sets.



Decoder for PFM system



Decoder for PTM system

The PFM and PTM systems have about the same reliability and simplicity. The PRR system is the most reliable and simple. It requires no adjustment in the airborne equipment and only two adjustments in the ground equipment. The PFM and PTM systems are more critical in operation.

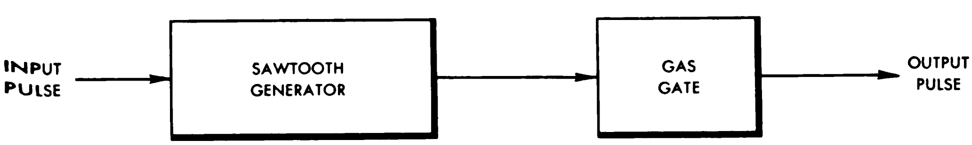
Pulse-Width Discriminator

A discriminator circuit used to pass pulses of greater width than a certain predetermined width, and reject all narrower pulses, is shown in the block diagram below.

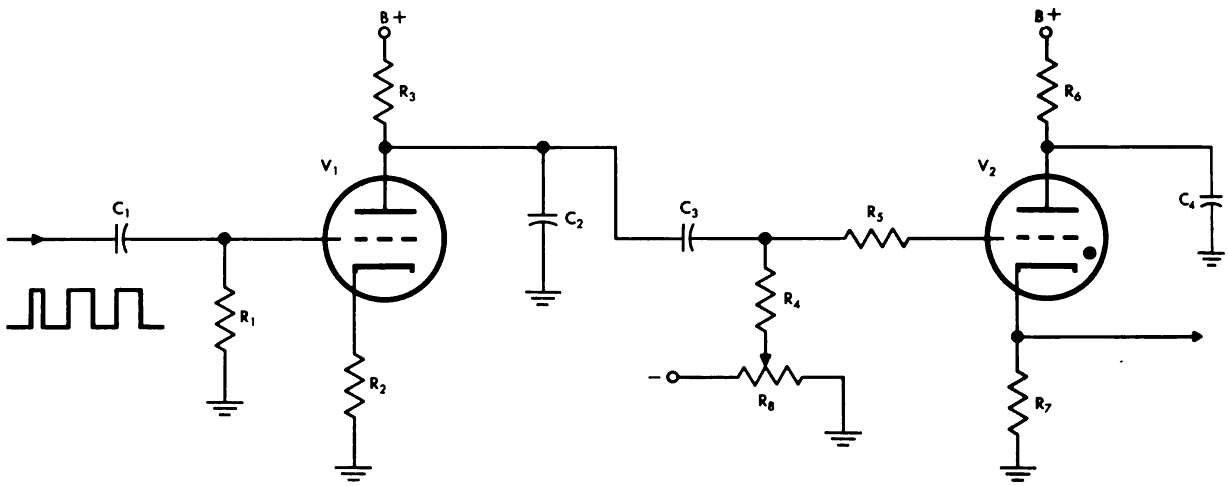
Preceding circuitry develops input pulses of constant amplitude but varying widths. There are output pulses from the decoder only if the input pulses exceed a certain preset value which is adjusted by a potentiometer setting. Noise and interfering pulses of narrower widths are rejected and produce no output.

The stages shown in the schematic drawing on the next page consists of a sawtooth generator (V_1) and a gas gate (V_2). Tube V_1 is normally conducting with a resulting low plate voltage. Negative-going input pulses of sufficient amplitude to cut off V_1 would normally allow the plate voltage to rise to "B+" and to develop an amplified and inverted version of the input signals in the plate circuit. The integrating condenser (C_2) in the plate circuit of V_1 develops a sawtooth output. It is important to realize that the amplitude of the sawtooth output increases with the increase in pulse width.

Tube V_2 is biased below cutoff by R_8 negative voltage. The setting of this potentiometer determines the amplitude of the positive-going sawtooth signal which allows V_2 to conduct. This gas tube ionizes when the grid signal sufficiently decreases the negative voltage on the grid of V_2. The tube V_2 will conduct through the output cathode resistor (R_7) until the plate voltage decreases to the point at which the tube deionizes. The output pulse from the cathode of V_2 is delayed with respect to the input pulse as shown on the next page. This delay is caused by the time required for the sawtooth signal to decrease the negative voltage on the grid of V_2 enough to allow conduction in the tube. Potentiometer tap R_8 can be adjusted for a smaller or greater negative voltage to require a different pulse width to trigger V_2.



Pulse-width decoder



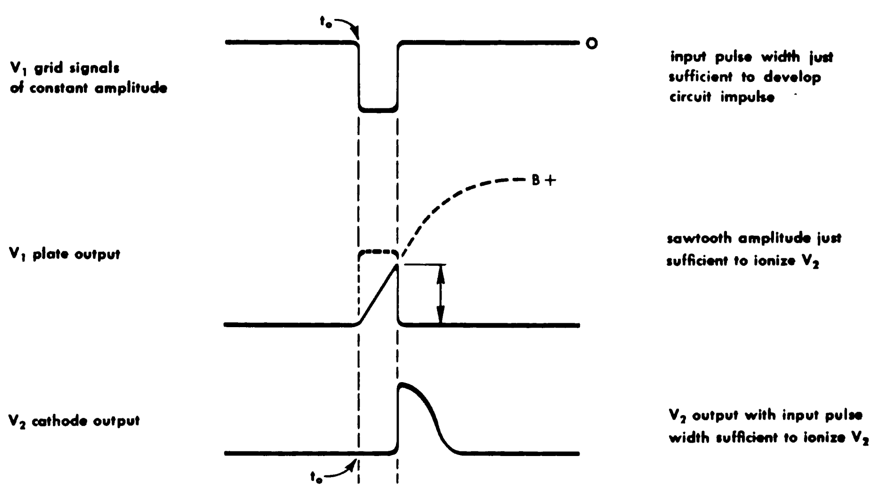
Sawtooth generator and gas gate

PULSE-WIDTH DISCRIMINATOR USING COINCIDENCE TUBE. The schematic on the following page shows a pulse-width discriminator circuit using a coincidence tube. In order to make the circuit operative, the pulse that is being selected must be changed into a sawtooth waveshape. These waveshapes are illustrated in the bottom figure on the next page as waveshapes "a" and "b." This change would be performed by a sawtooth generator stage prior to the input of this circuit.

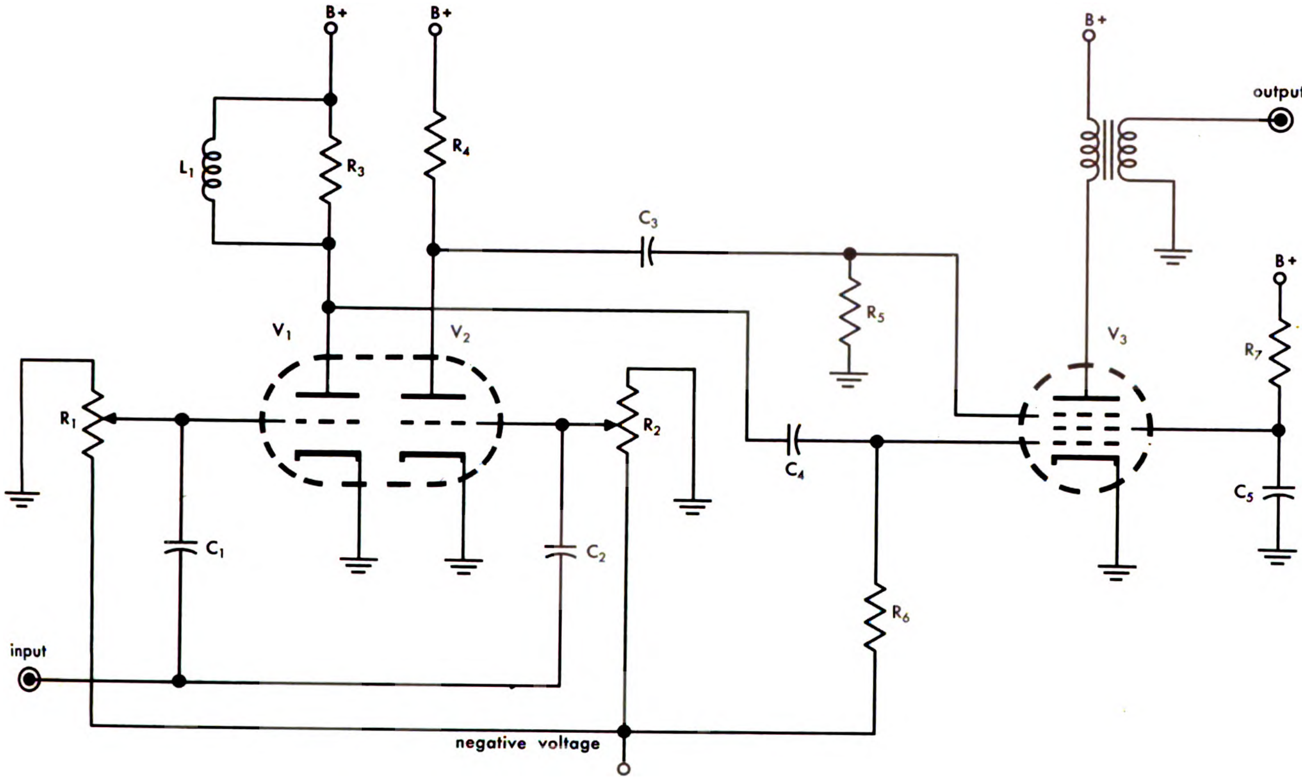
Note that all of the tubes are biased below cutoff by means of a negative-bias supply voltage. The bias on tubes V_1 and V_2 can be varied by moving the wiper arms on potentiometers R_1 and R_2 respectively.

Assume that this circuit is to be adjusted to pass pulses of 0.1 to 1.0 microsecond duration and reject all others. This means that 0.9 is the minimum and 1.1 is the maximum pulse width to be passed. V_1 is the tube used to determine the minimum pulse width. The bias on V_1 is adjusted so that V_1 will begin conducting when the signal voltage (E_1) on the control grid of V_1 is equivalent to the value at 0.9 microsecond after the start of the sawtooth wave. (See waveshape b.)

You can see that a surge of current through the RF choke L_1 would be shunted by R_3, causing a damped oscillation or ringing effect at the natural frequency of the choke plus the capacity of the circuit. The small resistor



Pulse-width decoder output waveshape with sufficient width input pulse



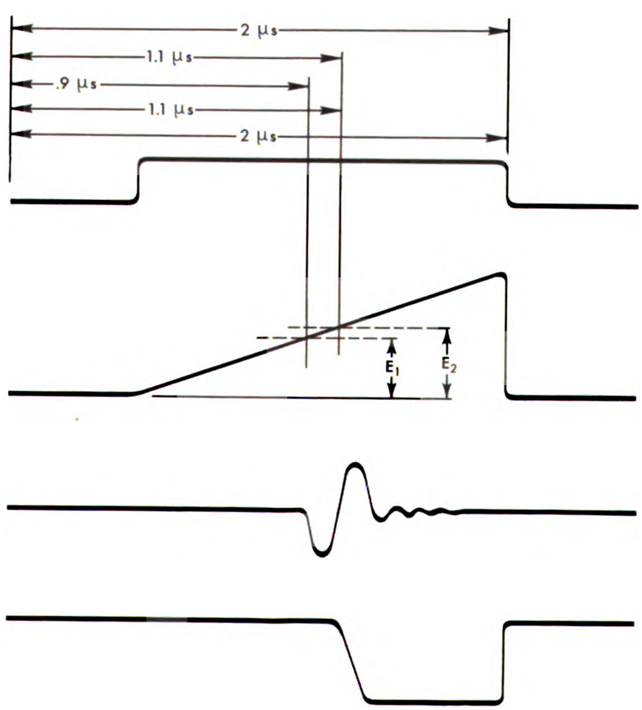
Pulse-width discriminator circuit using conicidence tube

a. pulse to be discriminated

b. sawtooth wave input produced from pulse above

c. signal at the plate of V_1 and control grid of V_3

d. signal at the plate of V_2 and suppressor grid of V_3



Waveshapes of a PWD which uses a coincidence tube

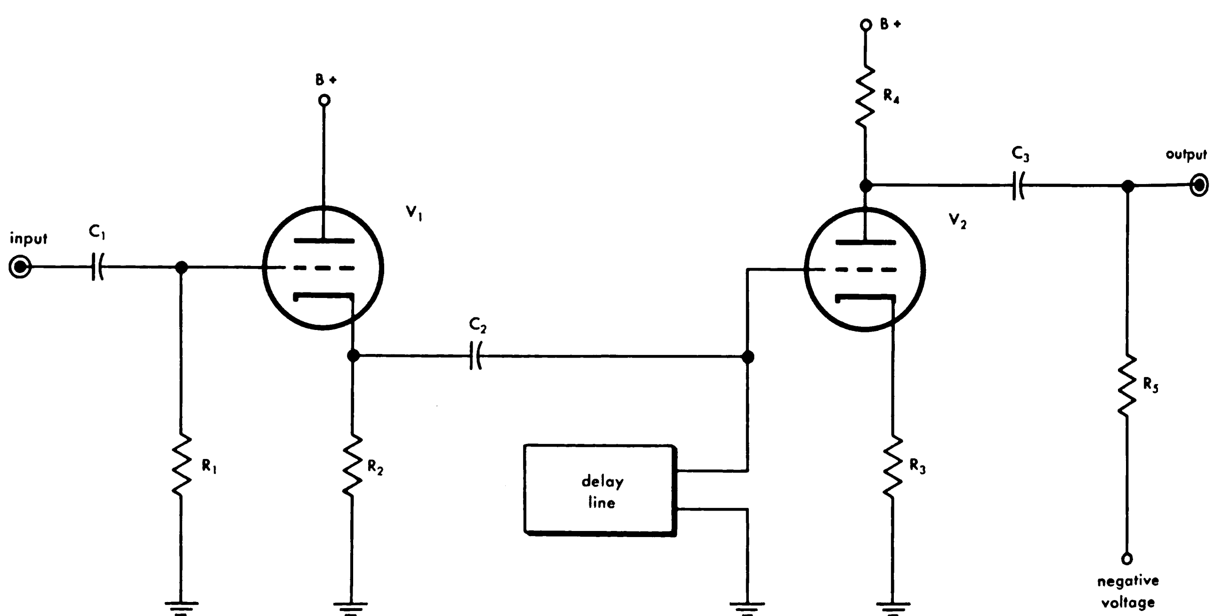
across the choke reduces the oscillation to only one complete cycle. (See waveshape c.) The signal on the control grid of V_3 is one sine wave, first negative and then positive, starting in coincidence with the time at which the negative-bias voltage is overcome by the positive sawtooth wave. You can see that if the pulse is too narrow to cause the sawtooth to reach an amplitude great enough to overcome this bias, V_1 will not be triggered and will thus discriminate against that particular pulse.

V_2 is being used as the maximum pulse-width discriminator. Again, as in the case of V_1, the negative bias has to be set by moving the wiper arm of R_2, but in this case the setting will be a different voltage. R_2 must be set so that V_2 conducts at the sawtooth voltage (E_2) equivalent to 1.1μs (see waveshape b). The plate load (R_4) is very large. This causes V_2 to be easily overdriven (saturated), and the signal on the plate approximates a square wave (see waveshape d). This signal is fed to the suppressor grid of V_3 and is large enough in amplitude to keep the tube from conducting, even though the control grid might be above cutoff voltage. If there is a pulse that is too wide, both V_1 and V_2 will conduct, and both signals will show in coincidence on the grids of V_3. The time relationship between the two signals, if they are present, is shown in waveshapes "c" and "d." V_3 is also biased below cutoff by means of the negative-bias supply. Because the sine wave first goes negative, the signal does not put the tube into conduction until the positive half-cycle is reached. If the pulse is too wide, the negative gate appears on the suppressor grid at the same time that the positive half-cycle of the sine wave appears on the control grid. The negative gate keeps current from flowing in the plate circuit while the positive half-cycle takes the control grid above cutoff. This means that the screen grid will draw current and there will be no output signal from V_3.

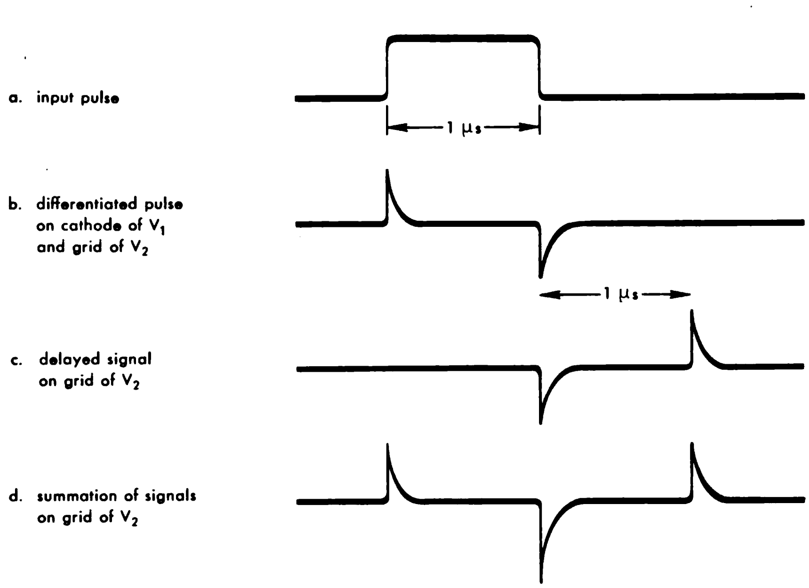
When a pulse of the desired pulse width comes in, V_1 conducts and V_2 will not conduct. The only signal to V_3 is the sine wave on the control grid, which produces one surge of current in the plate circuit during the positive half-cycle of the sine wave.

If you should desire to reshape this pulse, you could use a blocking oscillator or a one-shot multivibrator to give you a square pulse.

Because of the critical operating points of the grids of V_1 and V_2, a stabilized negative bias supply is required to give accurate pulse-width discrimination.



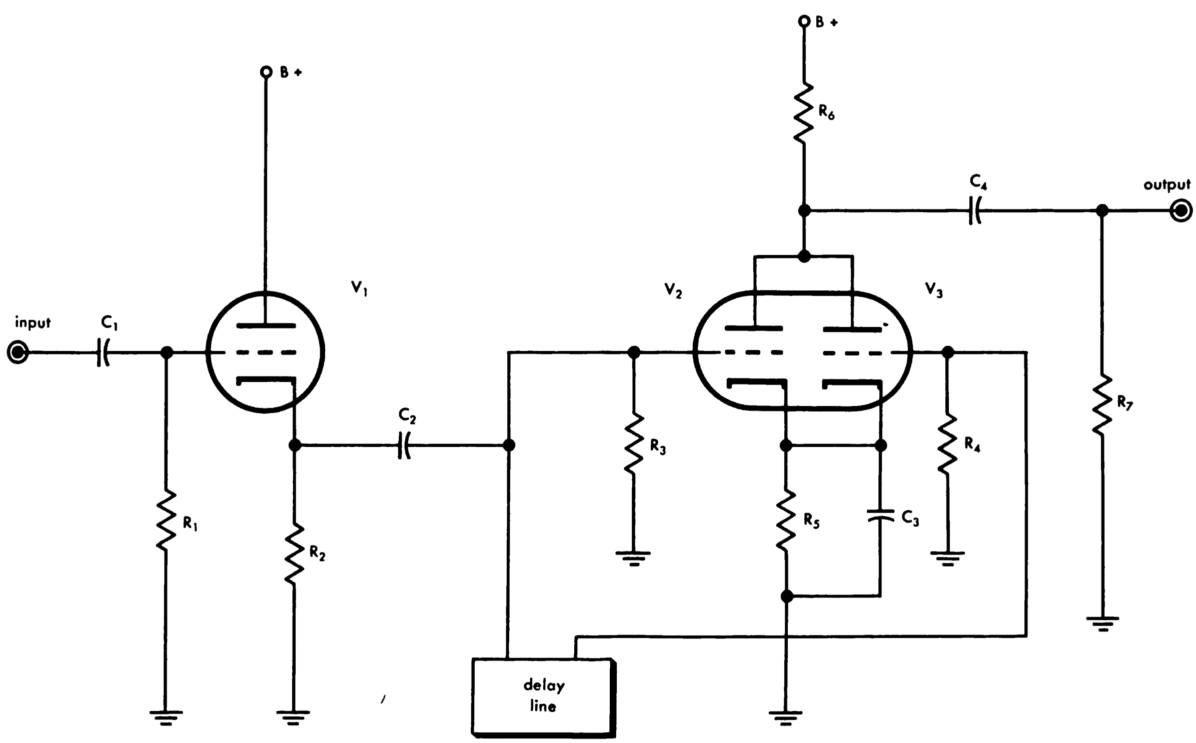
Pulse-width discriminator using a delay line



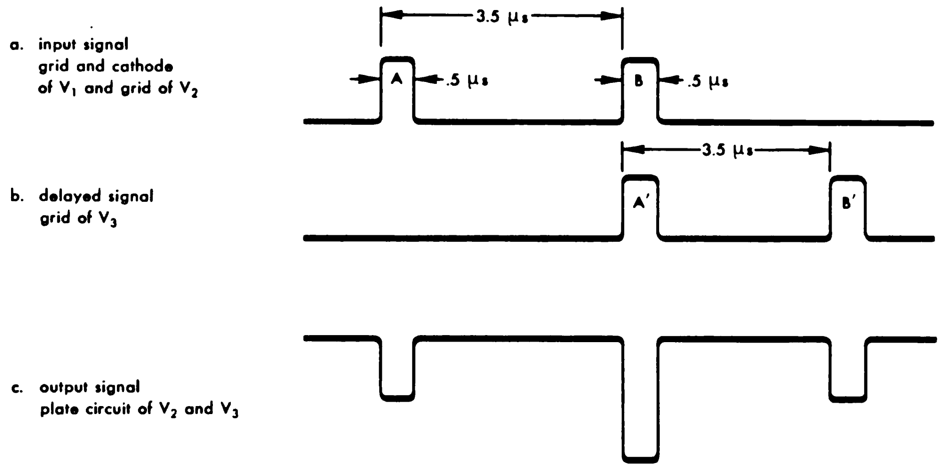
Waveshapes of a PWD which uses a delay line

PWD USING DELAY LINE. Another pulse-width discriminator employs the use of a delay line, as shown on the left. As before, let's assume that the circuit is set to pass 1.0 μs pulses and no others.

C_1 and R_1 differentiate the input pulse forming a positive peak on the leading edge and a negative peak on the trailing edge. Note waveshape "a" of the illustration above. These two pips (the positive and negative peaks) are



Pulse-spacing discriminator-type I



Waveshapes of a PSD-type I

separated by a time lapse equal to the duration of the pulse. Tube V_1 acts as a cathode follower, and the pips appear across R_2 and are coupled to a delay line and the control grid of V_2 by condensor C_2.

The delay line is grounded on the other end so that a negative reflection of the positive pip appears 1.0μs after the pip appears at the grid of V_2. (The positive pip travels for .05μs to the shorted end of the delay line where it is inverted, and in .05μs it travels back to the input end of the line.) If the reflection coincides with the negative pip from the trailing edge of the input pulse (see waveshape c), a large-amplitude negative pip results from their summation (waveshape d).

Tube V_2 inverts this signal, and the signal is then coupled to the next stage by means of C_3 and R_5. R_5 is connected to a negative voltage which biases the following stage far enough below cutoff.

The positive signals at the grid of V_2 would result in negative pips at the plate of V_2. These could not make the following stage conduct. Therefore, the negative pips at the grid of V_2 are added to each other, resulting in positive pips at the plate of V_2. The pips are then of sufficient amplitude to make the following stage conduct.

This following stage could be part of a pulse-forming circuit which is heavily biased to eliminate reaction to all but the large signal developed above.

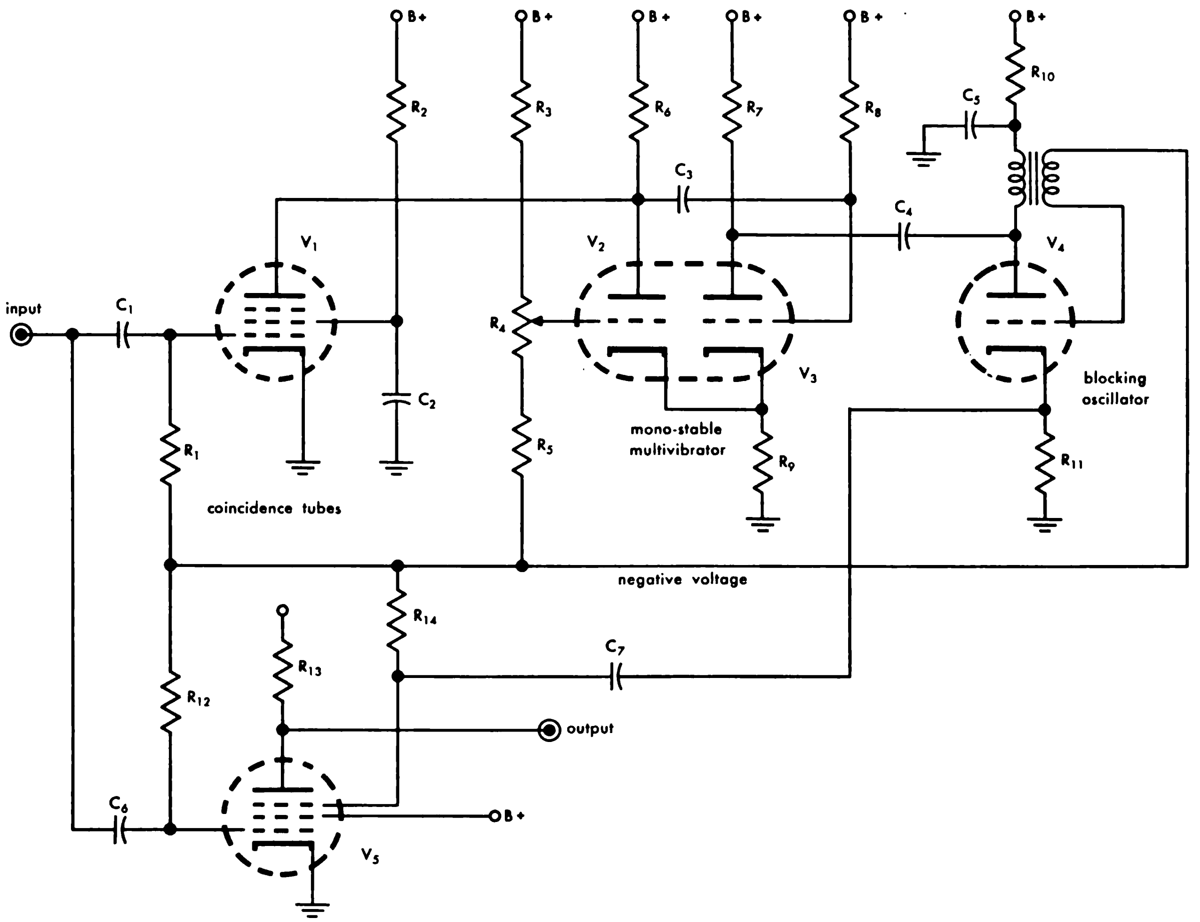
Pulse-Spacing Discriminators

Two types of pulse-spacing discriminators are discussed here. Basically, the first type consists of a delay line and a coincidence circuit, and the second one consists of a multivibrator, a blocking oscillator, and a coincidence tube. The applications of these two circuits will be discussed after the circuitry has been examined.

The pulse-spacing discriminator shown on the preceding page makes use of a delay line and a coincidence circuit. The cathode follower stage, tube V_1, is necessary to maintain a constant impedance to the delay line. The leading edges of pulses to be passed are spaced 3.5 microseconds apart, and the pulses themselves are 0.5 microseconds wide. Note waveshape "a" in the illustration above.

The signal is fed to the control grid of V_2, the first tube in the coincidence circuit, and it also is fed to the delay line. The delay line is tapped in such a manner that the signal is delayed by an amount exactly equal to the spacing of the two pulses (waveshape b) and then fed to the control grid of V_3, the second tube in the coincidence circuit.

Both of these signals change the current flowing through the plate load resistor R_4. The first pulse (A) on the grid of V_2 causes the plate voltage to drop, giving a negative pulse at the plate. The second pulse (B) on the grid of V_2 also causes a negative pulse at



Pulse-spacing discriminator-type II

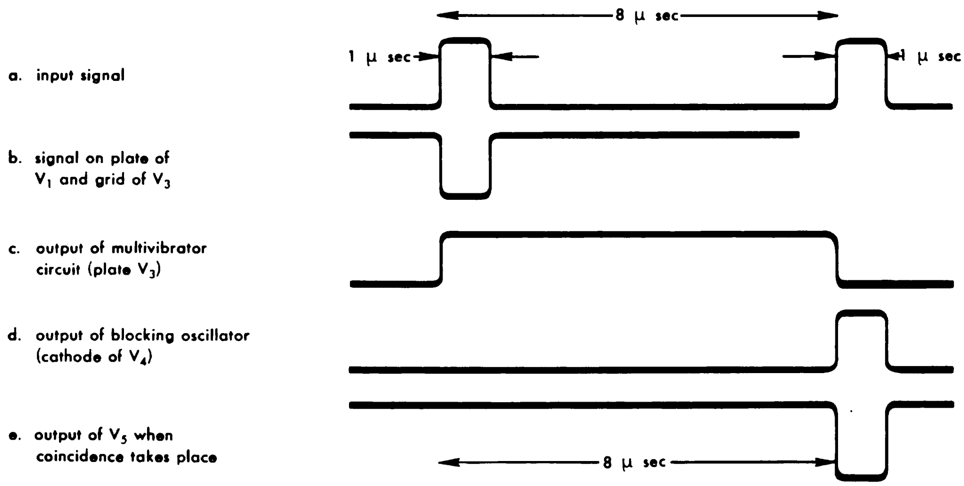
the plate, but at exactly the same time the first pulse (A′) of the delayed signal appears on the grid of V_2, causing a further drop in plate voltage (waveshape c). When the second pulse (B′) of the delayed signal appears on the grid of V_3, it causes a negative pulse at the plate. This pulse is the same amplitude as the first pulse "A" on the grid of V_2.

The accuracy of this circuit is good, being approximately ± .05 microsecond.

This signal on the plate of the coincidence circuit can be used to trigger a number of different circuits to indicate that pulses of that spacing have been received at the input. The signal could be used as a positive trigger by inverting it through an amplifier. Naturally, any circuit triggered by the signal would have to be biased so that only a signal composed of two pulses appearing *simultaneously* on the grids of V_2 and V_3 would have sufficient amplitude to trigger that circuit. Therefore a pulse applied to only one grid (V_2 or V_3) will not yield an output to trigger the next circuit. The pulses which are applied to the discriminator must be uniform in amplitude. This uniformity of amplitude can be accomplished with a limiter stage.

The second type of a pulse-spacing discriminator is shown in the schematic above. Tubes V_1 and V_5 are coincidence tubes, and tubes V_2 and V_3 compose the multivibrator circuit which is designed to pass pulses which are spaced 8 microseconds apart. The accuracy of this circuit (± 9.5 microseconds) is not nearly as good as the other pulse-spacing discriminator.

The output of the multivibrator is taken off the plate of V_3 and is a positive gate (waveshape "c"). The width or duration of the gate can be varied by potentiometer R_4.



Waveshapes of a PSD-type II

The gate is differentiated by coupling capacitor C_4 and the low-impedance winding of L_1, the blocking oscillator transformer. The differentiation of the trailing edge results in a negative pip which triggers the blocking oscillator. The signal, a square pulse (waveshape "d" above) is taken off the cathode of V_4 and coupled to the suppressor grid of V_5. At the same instant the second pulse of the input signal is appearing on the control grid of V_5. V_5 conducts and a negative pip appears on the plate (waveshape "e"). This negative pip is the output signal and indicates that two pulses of the proper spacing have been received. Both the control and suppressor grids of V_5 are biased below cutoff. There will be no current in the plate circuit of V_5 unless there is *coincidence* between a signal on the control grid and a signal on the suppressor grid.

Pulse-Recurrence Frequency Discriminator

The circuit of a pulse-recurrence frequency discriminator functions as a decoder by accepting only a limited band of pulse-recurrence frequencies (PRF) as synchronizing signals. This band of PRF can be preset by a potentiometer adjustment. Pulses at recurrence frequencies outside of this limited band will not produce a synchronized output from the discriminator and can be effectively gated out of succeeding circuitry by delay-line and coincidence-tube combinations already discussed.

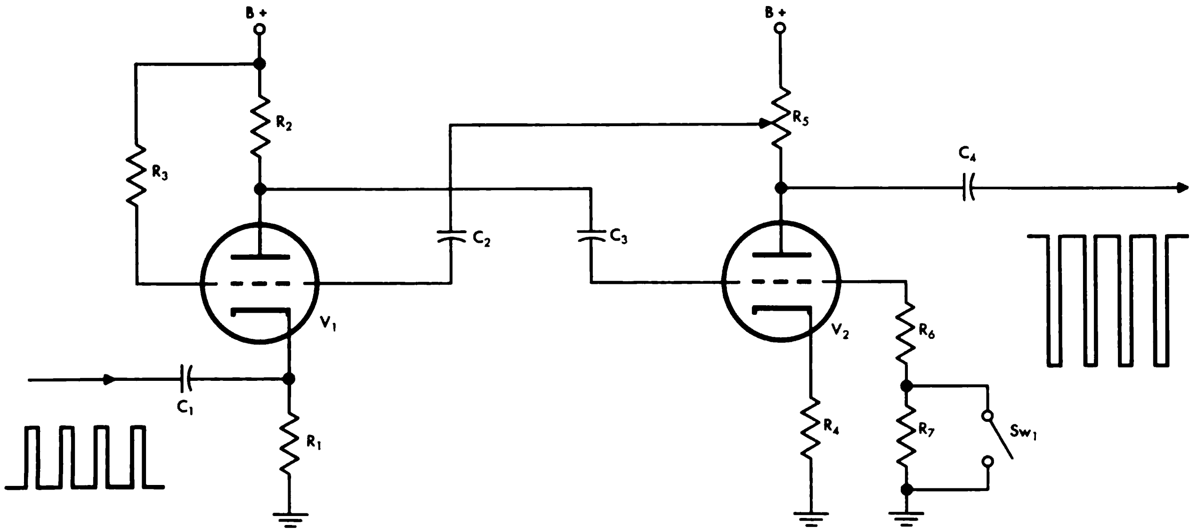
Tubes V_1 and V_2 of the diagram on the right and their circuit components comprise a free running multivibrator. Positive-going input pulses within the synchronizing range produce negative-going output pulses of the same recurrence frequency. The discriminator output is used to trigger the following circuit, such as a blocking oscillator.

Potentiometer R_5 setting in the plate circuits of V_2 determines the band of PRF desired to synchronize the discriminator. As the tap on R_5 is raised toward B+, thus decreasing the resistance through which C_2 discharges, the band of PRF that will synchronize the circuit is raised in frequency. Choice of component sizes determines the range of the frequencies. At the higher-frequency bandwidth setting, the range of synchronizing frequencies is also increased.

The switch in the cathode circuit of V_2, when closed, decreases the bandwidth from two to three times. Closing the switch makes the discriminator more selective to variations in the pulse-recurrence frequencies.

Phase Comparator

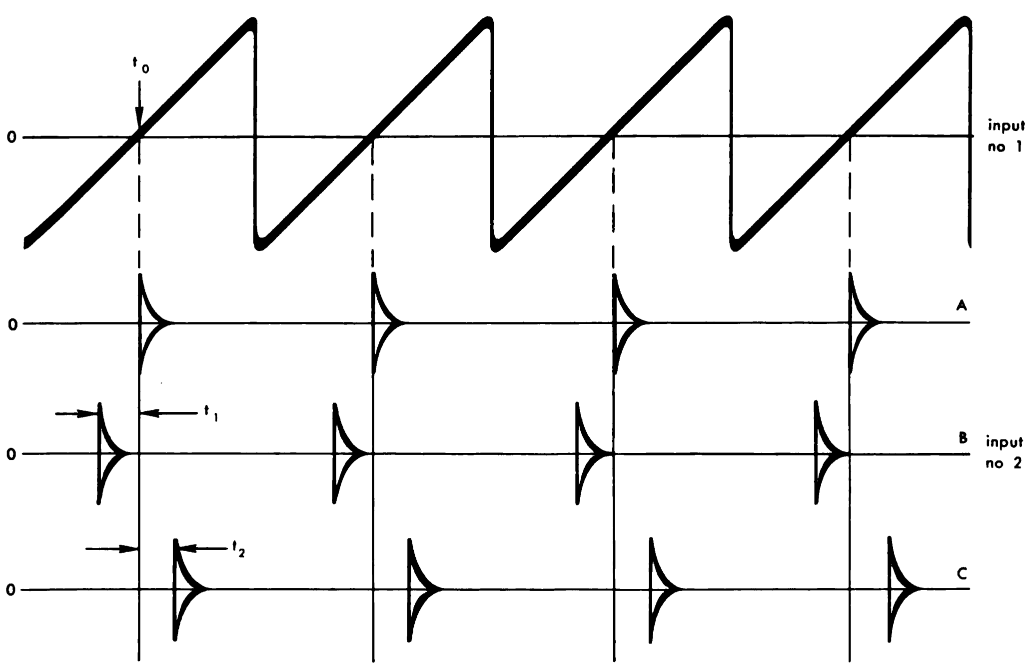
A phase comparator detects deviations from a preselected value of time difference between two sets of input pulses. There is no output from the circuit when the input pulse trains are of the desired time difference. The output is negative if the time difference between the pulses is greater than the selected value. The output is positive for a time differ-



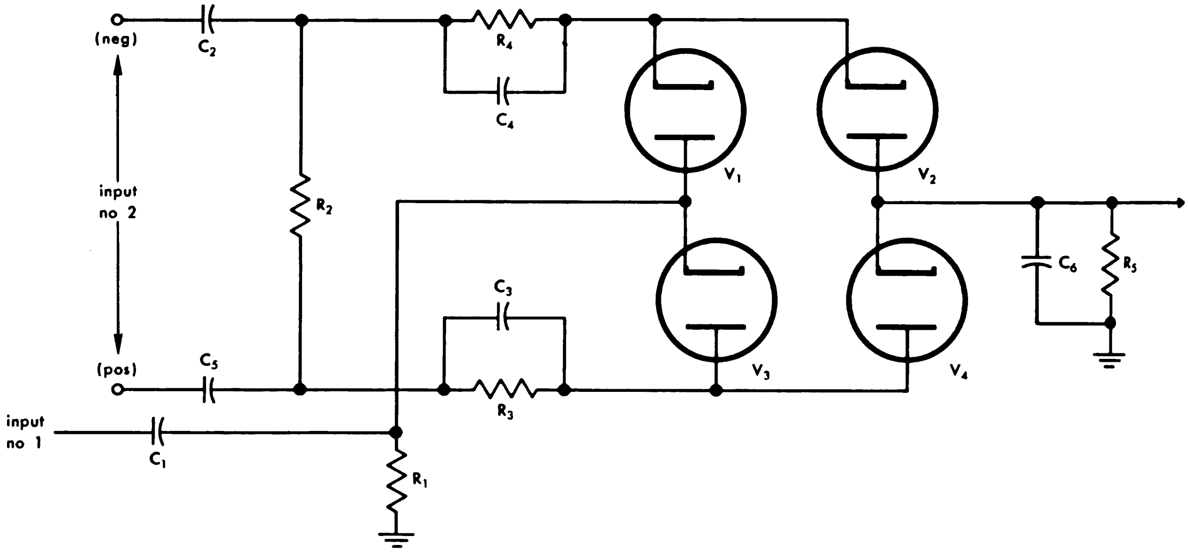
PRF discriminator

ence smaller than the selected value. The magnitude of the output voltage is proportional to the magnitude of deviation. The polarity of the output represents the direction of deviation.

A sawtooth voltage, input No. 1 in the accompanying graph, is developed by a pulse train of constant pulse-recurrence frequency. The middle of each sawtooth, which is of zero voltage value, is considered as zero time reference (t_o). Input No. 2 consists of pulses 180° out of phase but with the same PRF as input No. 1. The pulses at input No. 2 are developed through transformers or paraphase



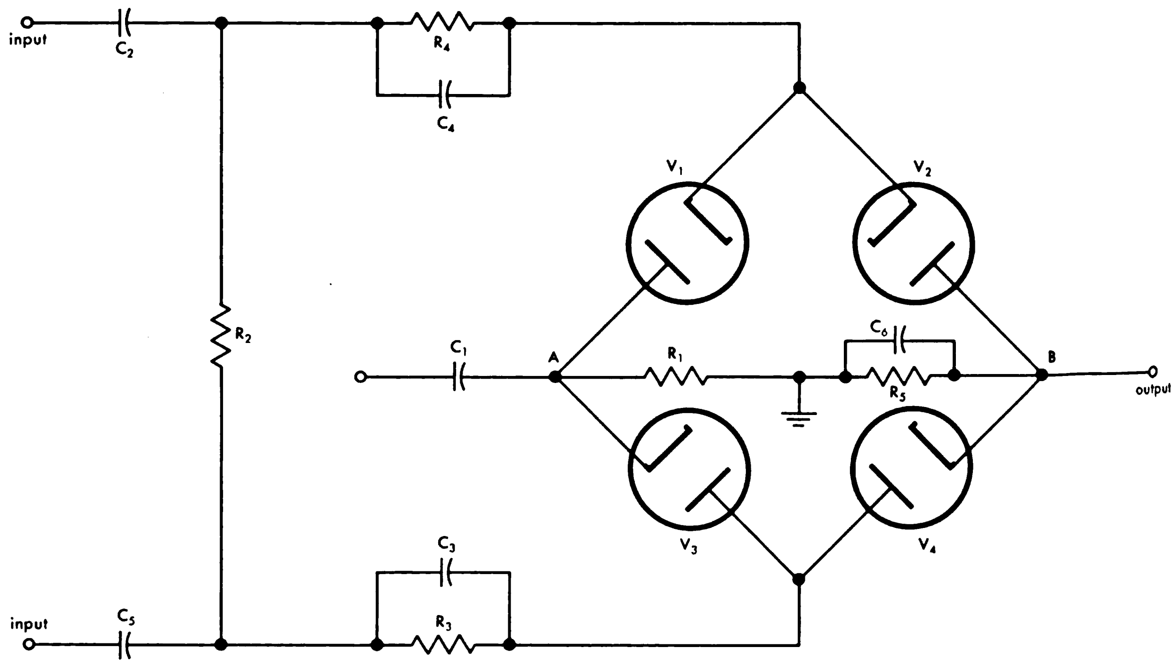
Pulse time difference relationships



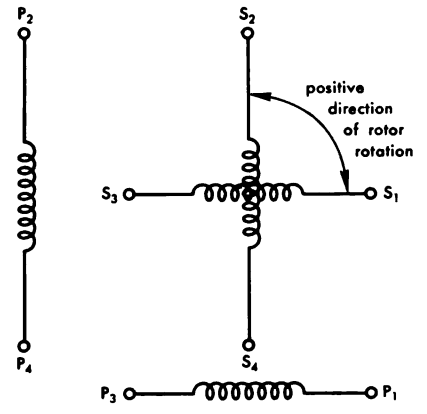
Phase comparator

amplifiers according to the deviation and are displaced a certain amount from the zero time reference. The combined effect of input No. 1 and input No. 2A results in zero output from the discriminator, representing an on-course condition.

A phase comparator circuit is compared to a bridge circuit in the schematic below. Without the sawtooth voltage input at terminal No. 1 of the previous schematic, the bridge circuit is similar to a balanced bridge with no output.



Bridge circuit representation of phase comparator



Resolver schematic with rotor in zero position

In this case, the input No. 2 pulses allow the four diodes to conduct equally and develop no voltage difference between points "A" and "B." With no voltage difference between these junction points, no current flows in R_1 and R_5 resistors, resulting in zero output voltage. This result is duplicated when the sawtooth voltage input at No. 1 and the pulse inputs at No. 2 are of zero time difference as shown in the case of input No. 2A of the previous graph.

Consider the time when the input No. 2B pulses are applied to the input terminals while the sawtooth input No. 1 is negative. Tube V_3 conducts more than V_1, starting current flow around the circuit R_3, R_2, R_4, and through V_2, the output resistor R_5, and then back to V_3 cathode through R_1. This current flow develops a negative voltage at the output terminal, junction of V_2-V_4, representing an off-course condition.

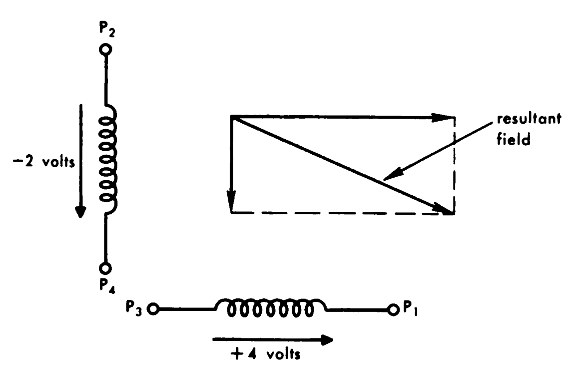
If the No. 2 input pulses are applied to the discriminator while No. 1 input is positive, as in the No. 2C case, current flow can be traced through V_1, R_1, R_5, V_4, R_3, R_2, and R_4. The output voltage is now positive, corresponding to an off-course condition in the opposite direction.

AUTOSYNS. Autosyns as reference units in control systems have been described in section III of the preceding chapter. The autosyn units in guidance computers are used to shift the phase of one or more voltages. These autosyn resolvers, as they are called, often function as coordinate converters.

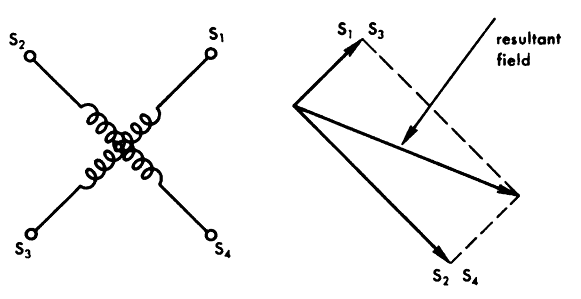
In some systems of reference, polar coordinates are converted to plane Cartesian coordinates, or plane Cartesian coordinates are converted to polar form.

Autosyn resolvers are essentially transformer arrangements with two primary or stator windings and two secondary or rotor windings. In the figure representing a resolver schematic with the rotor in the zero position, voltage applied to the P_1-P_3 winding induces maximum voltage of the same phase in the S_1-S_3 secondary winding. Voltage applied to the P_2-P_4 primary induces maximum voltage of the same phase in the S_2-S_4 secondary. With a small positive rotation from zero position, the voltage induced in S_1-S_3 is in phase with the P_2-P_4 excitation, and the voltage induced in S_2-S_4 is 180° out of phase with the P_1-P_3 excitation.

The output of each resolver secondary depends on the primary voltages and the rotor position. In the computing units, positive voltages representing positive quantities are



Resultant field of primary windings



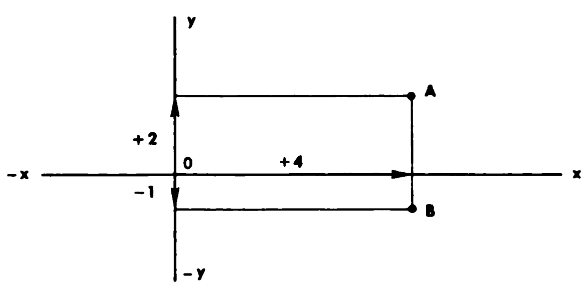
Secondary voltages

in phase with some reference voltage. Negative voltages representing negative quantities are 180° out of phase with this reference voltage.

The lower left figure on the preceding page shows the resultant field produced by the vector addition of the fields of the primary windings. These fields induce voltages into the secondary windings as shown in the last figure on the preceding page.

The action described in the preceding paragraphs is used in converting from one set of coordinates to another. A brief review of fundamental polar plane Cartesian coordinate properties follows.

In figuring these coordinates, a reference point in a plane and a reference direction are chosen. The position of any other point in the plane can be located by two numbers called coordinates. In Cartesian coordinates, both of the "x" and "y" coordinate numbers represent distances. In polar coordinates, one coordinate corresponds to a distance and the other to a direction.

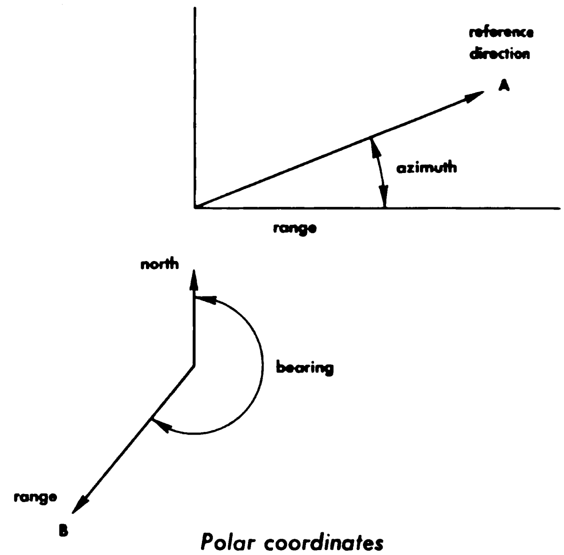


Plane Cartesian coordinates

In the illustration above, point "A" is represented by +4, +2 coordinates, and point "B" coordinates are +4, −1 from the origin 0.

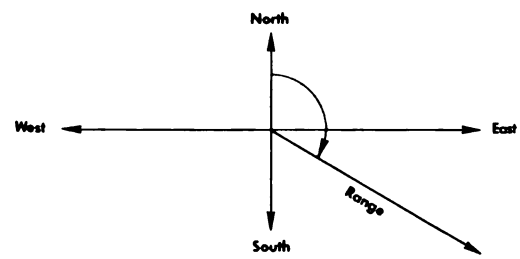
In the figure, following, the distances of the points "A" and "B" from the origin is the range, which is always positive. As shown, when the direction angle is measured from north in a clockwise rotation, the angle is called a bearing. When the angle is measured to a point from some direction other than north, it is called the azimuth.

Polar coordinates can be developed from a set of plane Cartesian coordinates by the method illustrated at middle right. Voltages



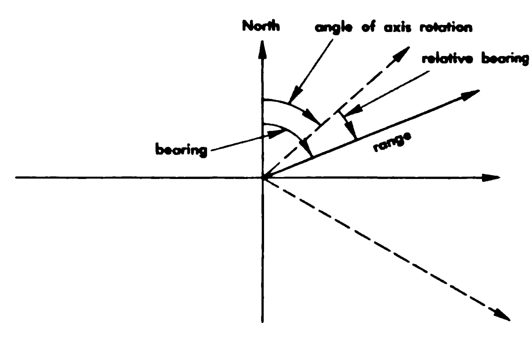
Polar coordinates

representing the east coordinate and the south (minus north) coordinate are vectorially added to form the polar coordinate, labeled range. In the resolver, the primary voltages may be considered the plane Cartesian coordinates. The resultant magnetic field develops voltages



Coordinate conversion

in the secondary windings that are the Cartesian vector components for a different set of Cartesian axes.

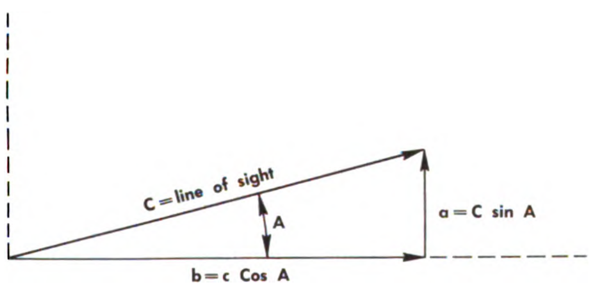


Rotating coordinate axes

The preceding figure illustrates the angles involved in rotating coordinate axes. First, the plane Cartesian coordinates are converted to polar coordinates as shown by the solid line, labeled *range*. The reference direction of this polar coordinate system is rotated through any desired angle. The new polar coordinates are then converted to plane Cartesian coordinates.

Autosyns are used in computer units as resolvers to produce voltages proportional to the sine or cosine of a measured angle. A voltage is applied to the stator winding which is to be resolved into components. This voltage amplitude represents a particular line-of-sight distance.

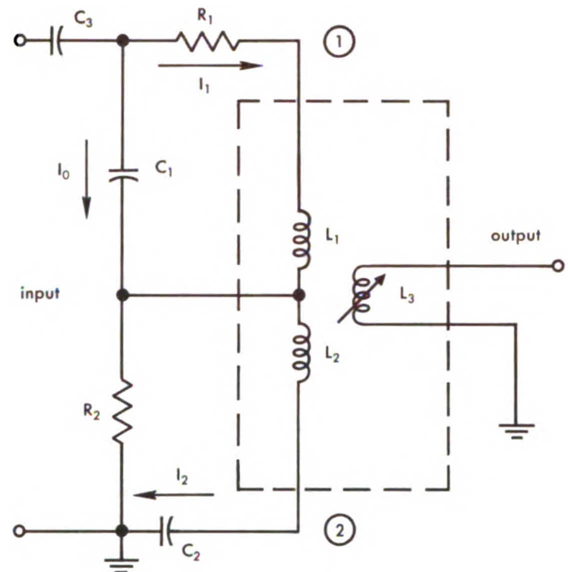
Voltage is induced in the rotor winding which is proportional to the angular displacement between stator and rotor. Assume an angle of value "A" as indicated in the below figure. Line "a" represents the value "C sin A," and line "b" represents "C cos A" from the trigonometric relationship.



Sine and cosine functions resolved graphically

An autosyn with its resistance-capacitance network is often referred to as a goniometer. Three types of goniometers are used. An uncompensated type without a counter is used in equipment to set up delays, synchronize pulses, and set up guidance pulses. Accurate time measurement of the pulses is not necessary, but the delays are variable and can be maintained stable.

A compensated goniometer with counters calibrated in microseconds is used in ground equipment to set up target information in a missile. There is also a motor-driven goniometer, rotated at a constant rate to simulate changing time differences.



Autosyn with associated circuitry

All these goniometers have the same circuitry, which is shown in the figure above. Coils L_1 and L_2 are wound physically at right angles to each other. Coil L_3 can be rotated through 360°.

ANALOG COMPUTERS OF MISSILE GUIDANCE SYSTEMS

Examples of mechanical devices that use analog computing principles include: the slide rule, the automobile speedometer, and the automobile differential gearing. In these examples, physical quantities are made to obey mathematical laws.

On the slide rule, logarithms of numbers are represented by lengths on the scales. Multiplication, division, raising to a power, or taking a root of a number is relatively accurate and rapid by operating on these quantities.

The automobile speedometer indicator is moved into position by the generation of electrical energy that is proportional to the speed of the drive shaft. This operation of indicating the rate of change of the automobile's position is one of differential calculus.

The automobile differential is a mechanical subtracting device. The amount of subtraction is determined by the gear ratio of the differential gears. Extra speed gained by one wheel

is subtracted from the other. This principle is used to add two quantities in mechanical differential analyzers.

The two quantities to be added are supplied to the ends of an adder. The sum is obtained by measuring the revolutions through which the body of the adder turns. Quantities to be subtracted can be represented by rotating the shaft in the reverse direction. Mechanical differential analyzers have been used for years to solve missile trajectories and problems in stability design.

Electronic differential analyzers are used in almost every phase of missile dynamic study to obtain information without constructing actual prototypes. A typical missile goes through several stages of development, such as: airframe design, design of the automatic control system, selection and design of the guidance system, analytical checks of performance, and actual checks of performance or flight-test analysis. The analog-type machines are able to study the flight trajectory without breaking it into separate phases, because they can handle continuous variables; and, in particular they can handle integration as a continuous process. One component common to the electronic analog computers is the *DC feedback amplifier*. This unit and its associated circuitry will provide our approach to the principles of analog computing.

First, let's consider how one of the more recent simulators or electronic analog computers is used in the missile field. This particular 4000-electron-tube machine simulates the characteristics of both missile and target. Through its use, the need for test firings of new missile designs is greatly reduced. Each of the machine's computing sections handles some particular phase of a complete missile system problem. The major sections are composed of the simulator, the target simulator computer, the aerodynamic computer, the guidance computer, and recording and display devices.

The guidance computer receives information from the simulator and the target simulator. Its output is fed to the aerodynamic computer.

The aerodynamic computer also receives information from the simulator in the form of missile altitude, velocity, and attitude. The guidance computer introduces missile fin-deflection information. The aerodynamic computer section determines what aerodynamic forces and torques should be delivered back to the missile simulator section.

The target simulator generates the target trajectory in the form of target positions in earth coordinates. Target maneuvers can be simulated by changing target speed, climb, and turn values.

Plotting units are normally used for missile and target trajectory plotting in both horizontal and vertical planes. Any 18 desired variables may be recorded by photoelectric recorders. There are two display devices: a model of the missiles assumes the altitude and fin deflection of the simulated missile; a trajectory model moves objects in three dimensions which represent the missile and the target.

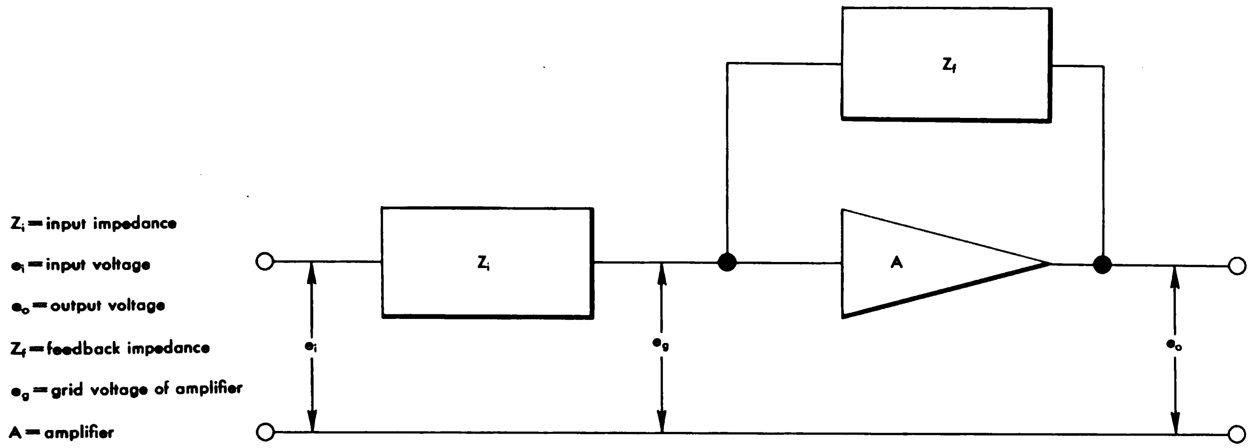
The following brief description of the inputs and outputs to the missile simulator computer illustrates the complexity of the circuits involved. Inputs in the form of voltages represent the aerodynamic forces along the roll, yaw, and pitch axes of the missile; the aerodynamic torques about these three axes; the initial position of the missile with respect to the earth's axes; the initial linear velocities along the three missile axes; and the initial attitude of the missile axes with relation to the earth's axes.

Outputs from the same missile simulator section are indications of the axes of the missile, initial linear velocities along the earth's axes, missile linear velocities along the missile axes, angular velocities about the missile axes, and the missile attitude in terms of the direction cosines of the missile with respect to the earth's axes.

Basic Operation of Feedback Amplifiers

One factor common to all electronic analog computers is the DC feedback amplifier. This basic component with its input and feedback networks will be described as it is used to perform the basic computer operations. One method of indicating the basic circuit is shown to the right above.

One of the most difficult design problems in DC amplifiers is the reduction of drift.



Basic feedback amplifier circuit

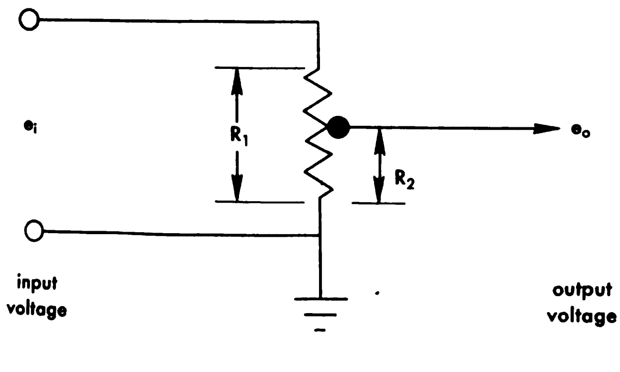
Special compensating circuits are used in combination with well-regulated power supplies to minimize this source of error in the output voltage. In some computer installations the vacuum tubes are aged to reduce the drift due to changes in cathode emission. Balancing the output voltage to zero for a zero value of input voltage is a regular procedure in some analog circuits.

Characteristics of most DC feedback amplifiers include high-input impedance, low-output impedance, and very high gain tubes.

The basic amplifier circuit can be used as a *sign changer*, because, when the input signal is applied to the grid, the output signal at the plate is reversed in polarity.

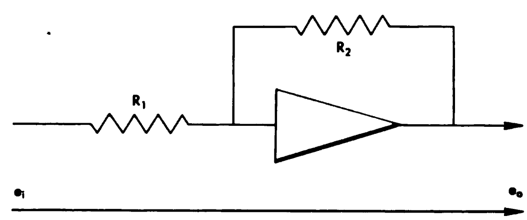
Multiplication of Voltages

One of the simplest methods of multiplying a voltage by a constant coefficient is to use a voltage divider or a potentiometer. In the diagram below, the proportion of input voltage e_i present at the output terminal e_o is equal to the voltage division of R_2/R_1.



Multiplication of voltage by a constant coefficient

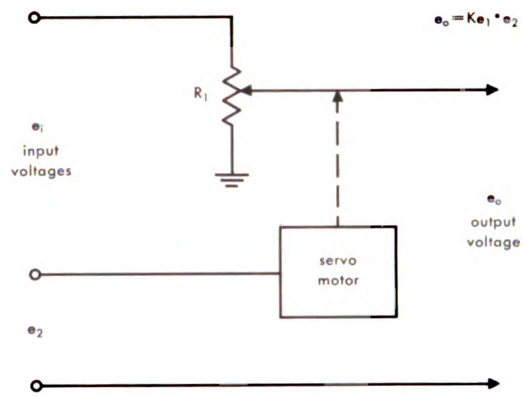
A DC amplifier also can be used with fixed resistors to multiply a variable voltage by a *constant coefficient.* In this case, the output voltage e_o is proportional to the ratio of the resistor values but with the sign changed to take into account the pulse inversion.



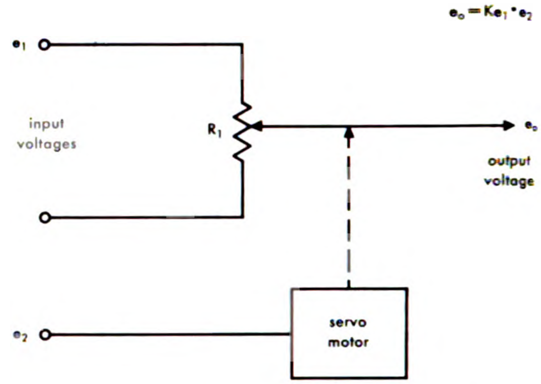
Multiplication of a variable voltage by a constant

Multiplication of two variables is commonly accomplished by using a servomechanism. Some type of servomechanism is used to position the potentiometer arm so that the amount of mechanical rotation is proportional to one of the input signals (e_2 in the schematic on the next page. Input signal e_1 is connected across the potentiometer. The output voltage e_o is proportional to the product of the two input voltages. The dotted line in the schematic represents a mechanical connection.

The connection shown in the accompanying figure allows multiplication of *two variables* for all combinations of signs. Values of e_2 represented by the position of the potentiometer arm can be positive or negative in sign. The computation speed is limited by the rate at which the motor driving the potentiometer arm can follow the changes in the voltages e_2.



Servo multiplier



A servo multiplier allows multiplication of two variables for all combinations of signs

Linear circuits alone cannot be used to multiply because the product of two quantities is not a linear type of expression. Multipliers must use devices in which gain depends on one or more of the applied voltages.

Electronic multipliers follow a much more rapid change of input signal than a servomechanism type. The multiplier unit is one of the basic forms of non-linear computing components. Linear amplifiers may be used to feed the variables into the multiplier.

The circuit shown below uses two sign-changing amplifiers. The circuit is connected so that both negative and positive values of input voltage e_1 are introduced to make the sign of the product correct.

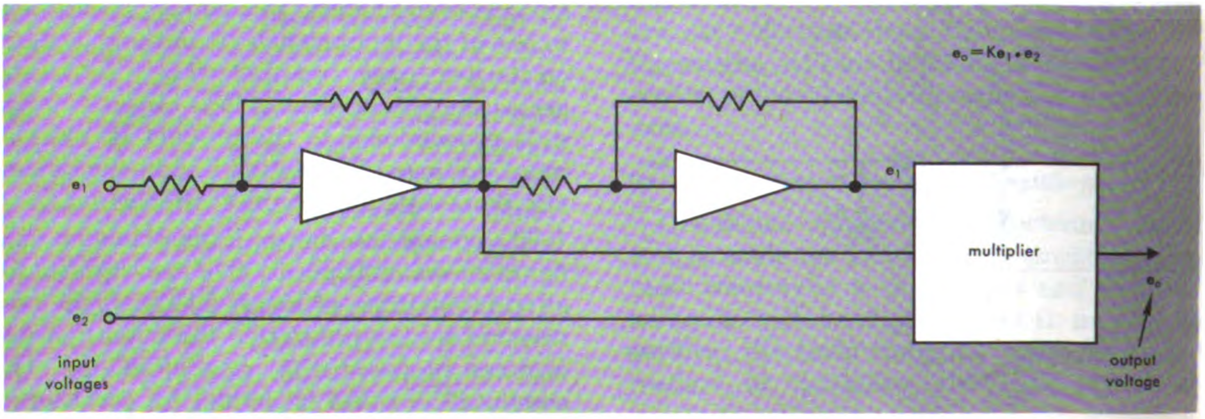
Division of Voltages

One method of dividing a variable voltage by a constant is to multiply the variable voltage by the reciprocal of the constant. When using this method, multiplying devices serve as dividers.

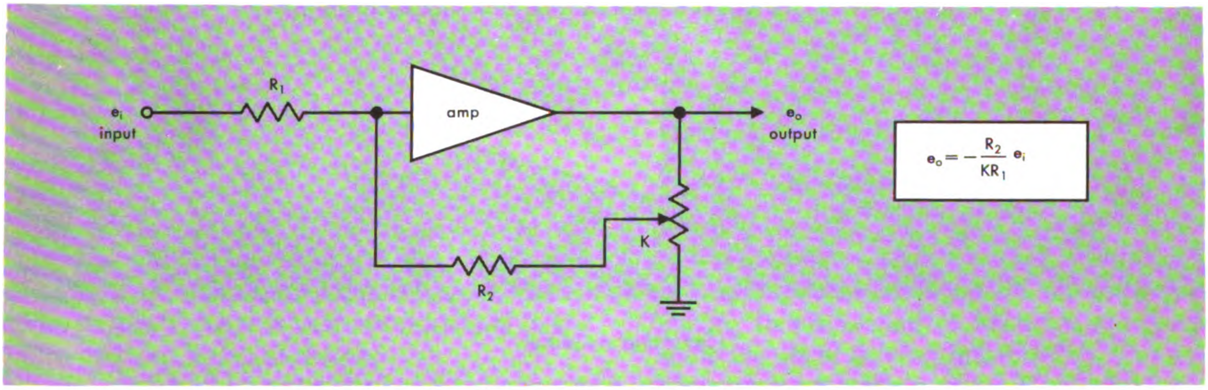
Another method of dividing a variable by a constant uses a high-gain negative feedback amplifier.

In the schematic top right, the constant "k" is determined by the setting of the feedback potentiometer arm in the amplifier output circuit. The loading effect of R_2 must be considered when the potentiometer is set. The constant must not approach zero. If it did reach zero, the result of the operation would not be accurate because the amplifier output would be limited.

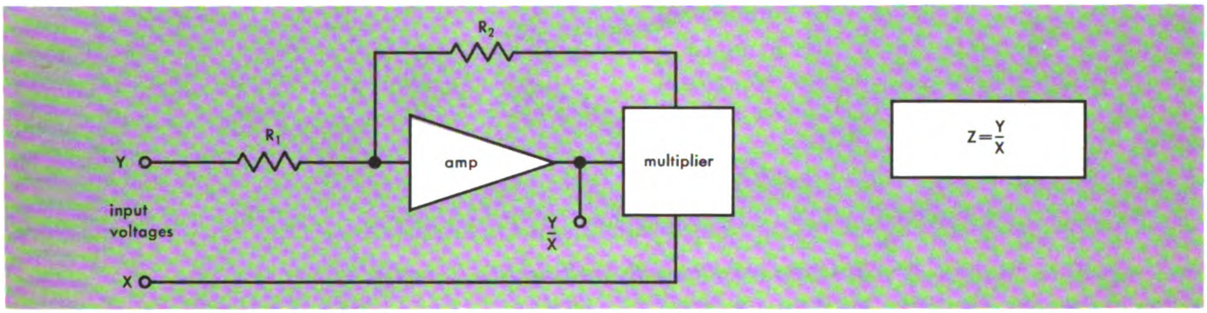
A high-gain amplifier is also used in a circuit which determines the quotient of *two variables.* An unknown quantity (Z) is assumed so that the product of "Z" times "X," one of the



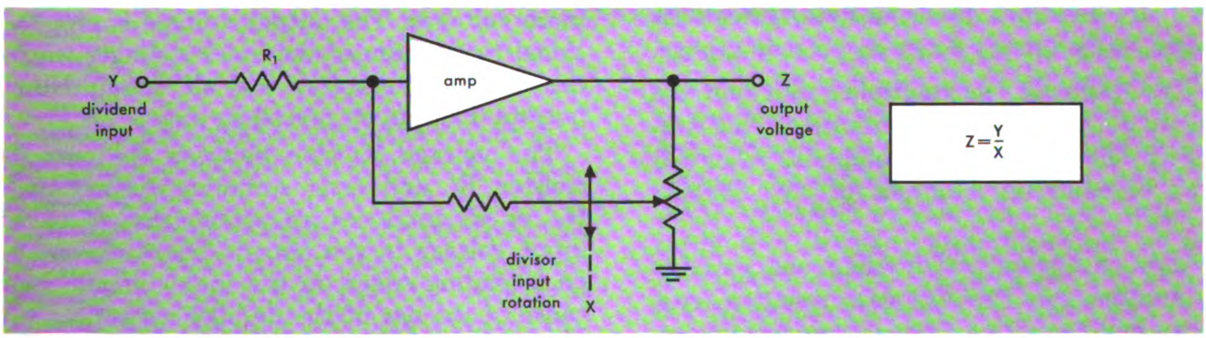
Multiplication of two variables using two sign-changing amplifiers



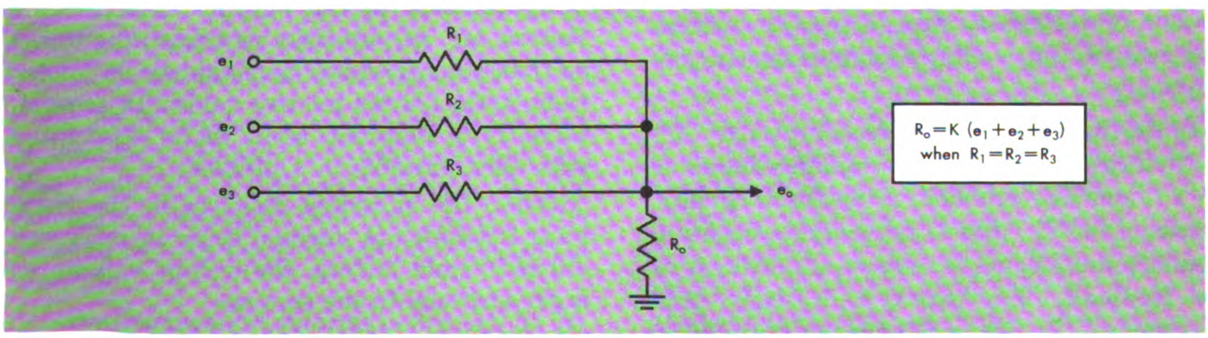
Division of a constant, using high-gain negative feedback amplifier



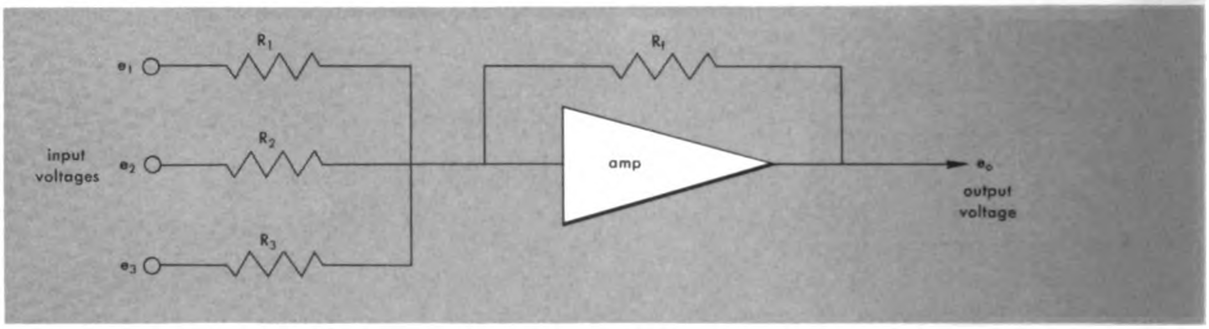
Division of one variable by another, using high-gain amplifier



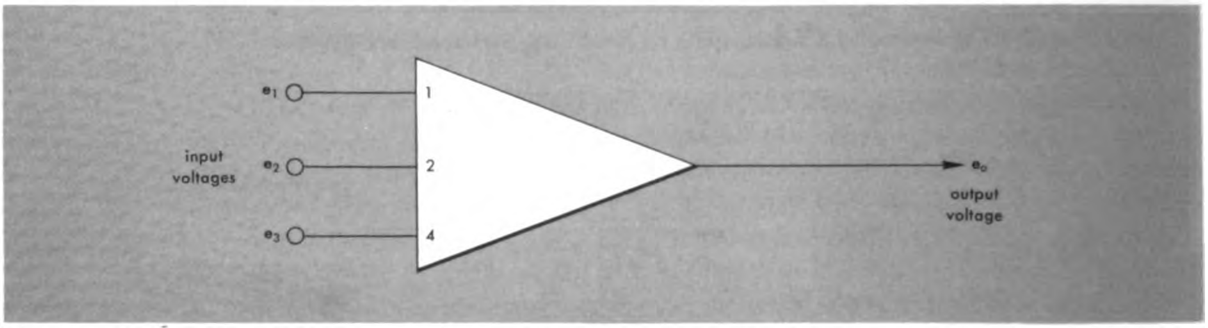
Division of one variable by another, using high-gain negative amplifier with feedback potentiometer



Summing resistor network



Summing amplifier



Summing amplifier with coefficients 1, 2, and 4

variables, is made equal to "Y," the other variable. The unknown "Z" and the variable "X" are fed into the multiplier circuit as shown in the second figure on preceding page.

The output of the multiplier unit and the negative "Y" value are summed in the amplifier. The output of the amplifier when fed back becomes the quotient of "Y" divided by "X."

Another method of dividing one variable by another uses the high-gain negative feedback amplifier with a feedback potentiometer. The dividend, or voltage to be divided, is supplied to the amplifier input terminal. The divisor, or voltage dividing, is used to position the arm of the feedback potentiometer mechanically.

In the third diagram on preceding page, "Y" is divided by "X." The amplifier input is "−Y" and the amount of feedback is proportional to "X." Since the gain of a negative feedback amplifier is inversely proportional to the feedback, the output is inversely proportional to "X," giving the result of dividing "Y" by "X."

Summation of Voltages

Summing networks are used to add two or more voltages. The output voltage is proportional to the sum of the input voltages when the summing resistors are all equal. Note bottom diagram on preceding page.

The voltage division between the output resistor R_o and each summing resistor R_1, R_2, or R_3 produces an attenuation of each input voltage in the circuit. There also may be a loading effect on the input circuits because of the use of the parallel network.

The value of the constant of proportionality "K" is determined by the ratio of the summing resistors to the output resistor. Scale factors are set up in the computing units so that quantities can be combined in operations in the correct relative values. In general, there is a different scale factor associated with each variable in a problem.

When the summing resistors in the summing network are unequal so as to represent different scale factors, the output voltage e_o is proportional to the sum of

$$\frac{e_1}{R_1}, \frac{e_2}{R_2}, \text{ and } \frac{e_3}{R_3}.$$

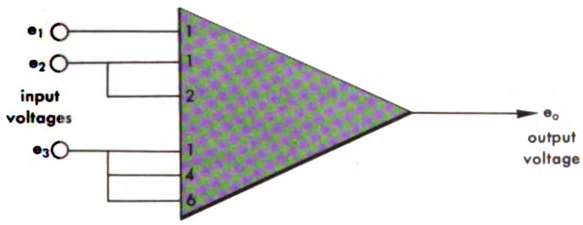
A summing amplifier also can be used to add several voltages.

The output (e_0) of the circuit shown at the left is inverted. The circuit can also be used to multiply by a constant.

If the summing resistors have the values of R_1=1 megohm, R_2=.50 megohm, and R_3=.25 megohm, the input voltage terminals correspond to coefficients 1, 2, and 4 respectively. The output voltage is then proportional to the sum of $1 \times e_1$, $2 \times e_2$, and $4 \times e_3$.

A summing amplifier with the coefficients described above is shown in block diagram form at the left.

A summing amplifier may be used to multiply several input voltages by constant coefficients. This is made possible by paralleling input circuits as shown below.



Parallel-input summing amplifier

In this particular example the output voltage is proportional to the sum of $1 \times e_1$, $3 \times e_2$, and $11 \times e_3$.

If the input voltages are positive, the output is negative. By means of potentiometers in the input circuits, this type of summing amplifier produces sums or differences of the input voltages with any combination of present coefficient values.

DIGITAL COMPUTORS OF MISSILE GUIDANCE SYSTEMS

In missile guidance, a tremendous amount of complex data needs to be evaluated quickly. Not only are computers essential to the guidance of the missile, but they are also quite necessary in the research and development of missiles.

Some telemetered information must be instantaneously evaluated for the range safety officer, while still other data must be evaluated so as to incorporate the findings of missile flights in future launchings.

A digital computer, as was stated earlier, is assigned a certain numbering system. It then performs a mathematical operation on these numbers (expressed in the form of digits) to yield an answer also expressed in digits.

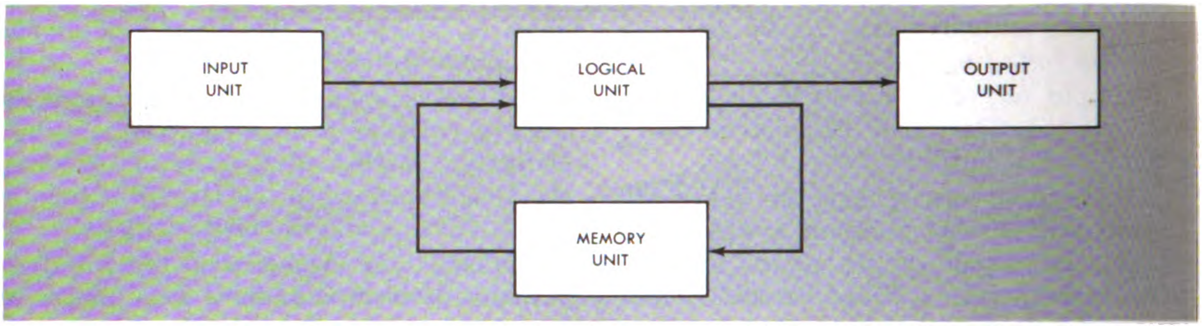
The first thing that we want to do in computing a problem is to have some means of storing this information. Man has always had some means of storing information. The most basic way of storing quantities is by use of the fingers. In this manner we assign various values to each finger; consequently, we have a reference for future use.

Most of the development on modern calculating or computing devices, as they are known today, was started in the late 1800s. These first machines were quite limited as to the operations they could perform and the speed at which calculations were made. The dial telephone was one of the first type of machines that could handle multiple operations successfully. The dial system is capable of handling several million telephones by successive relay operations according to assigned digits of the phone numbers.

Another modern development of calculating machines is that developed by the International Business Machines Corporation (IBM). A predecessor to the IBM machine was developed by Herman Hollerith in 1890. He made extensive experiments with punched cards that would permit electrical contact to be made during the presence of a hole in the cards. The device that developed from these experiments counted or added, singly or in combination, as desired. This punch-card technique is still a major component in our up-to-date computers.

To have a computing device that is capable of taking in a set of conditions (a problem) and determining the result of these conditions, it is necessary that it be capable of some of the traits which we, as humans, possess.

As you know, the computer must be capable of *storing* information. This storing of information is commonly referred to as the *memory unit* of the computer. Here, numbers and instructions are retained until the unit is called upon to use the information to com-



Functional diagram of digital computer

plete the problem. As you will see later in the discussion, the memory unit may be capable of retaining the information for long or short periods of time depending upon the type of device that is employed for this purpose. The memory unit can be likened to our brain; and it thus needs some means of transferring the information to other parts of the system. These links or channels by which the information is passed from one unit to another can be called the *data transmission system.*

Now that we have the material and a means of transporting it from one point to another, there must be a unit that will take care of the necessary arithmetical operations on the problem. This unit is referred to as the *logical unit.* The logical unit adds, subtracts, multiplies, and divides; it takes square root, attaches the proper sign, and tells what a quantity is equal to.

A *control system* is the next item that must be included in the computer. The control system tells the machine when to perform a certain operation. It also governs the type of operation that is to be performed; that is, it governs whether the operation is to be addition, multiplication, or any other arithmetical calculation.

Of course, there must be some provision made for the reception of data and the supply of answers in the machine. These units, representative of their functions, are called *input* and *output* units respectively.

Finally, there are motors and electricity that provide for the energy to operate the system.

Although a digital computer is versatile in problem handling, there are some "traits" it does not possess. It lacks intuition, which means that every problem must be presented to the system before it will operate properly. It does not possess the ability to make guesses or draw conclusions; therefore, it only gives answers according to the accuracy of the data presented. The computer is not capable of rearranging the sequence of operations or developing any of its own instructions; therefore, it cannot make any erroneous changes in the "programming" of the material. And it cannot form any situations that are not pertinent to the problem as given in the problem's original form.

Since humans have intuition, you might think it would be a valuable asset in machine problem handling. But by not having these characteristics, the mechanical brain can only deal with *facts* and deliver answers based on these facts. No extraneous matters intrude upon its straightline "thinking."

We will use the functional diagram above for the purpose of taking up the various components that may be used in the units of a digital computer.

Before proceeding with the discussion, it should first be realized that the components that carry out the various functions in the computer may not be physically distinct. The same components or types of components may be used for more than one function. For example, a component that makes arithmetical operations may also be used as a memory unit.

Binary Numbers

In a majority of the digital computers a binary system of coding is used. If a question is asked of a machine, the smallest unit that may be used would be one to represent the truth of the statement. The answer may be in the form of a *yes* or *no* (a check mark or no check mark) or, as in the case of electronic instruments, an *on* or *off* (an impulse or no impulse). In short, there are just two conditions that need to be considered to determine whether a statement is *true* or *false*.

In the binary digital computers, use is made of this fact by employing what is known as a *binary representation of numbers*. The twofold difference of representing information makes it possible to use devices that represent only the *go* or *no go* conditions. For example, a relay in a *closed* position could represent a *go* condition, while the same relay in an *open* position would represent the *no go* condition. The punch-card system used in computers lends itself well to this system. In this case, the presence of a hole would represent the *go* (yes) position and the absence of a hole would indicate the *no go* (no) condition. These two conditions are indicated in the binary numbering system by the two digits "0" and "1" respectively.

We are all familiar with the everyday usage of the scale-of-ten (decimal) representation of numbers. Let's take the number 8501 and determine how it is made up by powers of ten. In this number we are working with what is called the *powers of ten*, that is, with 10, 100, 1000, 10,000 — or 10 multiplied by itself as many times as desired. The number 8501 indicates:

$$(8\times1000)+(5\times100)+(0\times10)+1, \text{ or}$$

$$(8\times10^3)+(5\times10^2)+(0\times10)+1$$

In the decimal representation of a number, the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9 are used. In the binary representation of a number, only two digits (in various combinations), 0 and 1, are used.

In the binary system, a sequence of 0's and 1's indicates the presence of the number 2 (two) raised to some power. For example, the digits 1101 indicate $(1\times2^3)+(1\times2^2)+(0\times2^1)+1$. This would equal 13 in the decimal system.

SCALE OF TWO

1101

$$1\times2^3+1\times2^2+0\times2+1$$

Decimal and binary representation of numbers

Shown below is a table of the first decimal numbers and their equivalent in the binary system.

DECIMAL	BINARY	DECIMAL	BINARY
0	0	13	1101
1	1	14	1110
2	10	15	1111
3	11	16	10000
4	100	17	10001
5	101	18	10010
6	110	19	10011
7	111	20	10100
8	1000	21	10101
9	1001	22	10110
10	1010	23	10111
11	1011	24	11000
12	1100	25	11001

Table of decimal and binary numbers

Examine the preceeding table for a moment. Consider the first four decimal numbers 0, 1, 2, and 3. To use the decimal system of numbering in a computing device, you would need to have four *registers* in order to record the information presented. Now, from the table, you can see that the binary system needs only two digits, the "0" and "1," to represent the first four numbers. In this case, four conditions of information can be represented by the use of two digits.

You now should be able to see where an ON-OFF electronic system fits into the binary pattern very well. Say that you wish to indicate the decimal number 3 in a computing function. This can be accomplished in the binary digital computer by representing it (from the table) as 11. The binary digits (11) would be represented in a relay system by both of the relays being in the closed (go) condition.

Once again referring to the table of decimal and binary numbers, you see that three digits of the binary system will enable you to represent eight pieces of information; with four digits, you can represent 16 pieces of information. By the time you have seven binary digits, you can represent 128 pieces of information.

Boolean Algebra

Before a computer can operate properly and satisfactorily employ some numbering system, such as the binary system, it must have the ability to:

1. Learn and remember data presented to it
2. Make a choice based on previous results
3. Make long chains of operations
4. Determine if the answer is correct
5. Determine when one problem is finished and when to start on another

In order for the computer to be capable of the foregoing items, the proper information must be given to it. From this point on it will determine the answer by what is known as logical truth.

Logical truth differs from *ordinary truth* in that we deal not only with facts but also in suppositions based on facts. For example, the statement "sugar dissolves in water" is an *ordinary truth*. Here, however, certain things are understood. The amount of sugar is much smaller than the amount of water; for if we tried to mix a whole bag of sugar with a teaspoonful of water, all of the sugar would not dissolve. The computer cannot operate by understanding such limiting conditions, unless these conditions are given to it so that it may form a *logical truth pattern* to work from.

A logical pattern may be of the following nature:

1. All people are animals.
2. All animals are mortals.
3. Therefore, all people are mortals.

Such statements as the above are called logical syllogisms. They are truth statements that result in a truth statement. It should be understood that all syllogisms may not be logical. For example, in the following statements you can see where statement 3 is not true even though statements 1 and 2 are:

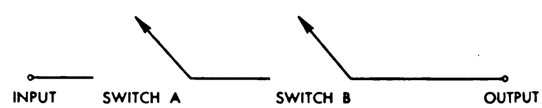
1. All animals have legs
2. All people have legs
3. Therefore, all animals are people

Many logical truth patterns are used by people in everyday life without their realizing that they are using them. Most of our simpler logical patterns are distinguished by words such as: *and*, *or*, *if*, *else*, *not*, *then*. In mathematical terms these logical patterns may be indicated by *plus*, *minus*, *times*, *divided by*, or combinations of these terms.

If a set of statements is made, it is possible to set up a table that will determine what is known as the *truth values* of these statements. In setting up this table, it is necessary to use some of the logical connectives, such as *and*, *or*, *if*, etc. In the end we would have a table that would list the truth or falsity of each of the conditions of the statements that we were considering.

It is possible to set up mathematical relations of the truth values of statements and perform mathematical operations on these statements. This study of mathematical logic is called the *algebra of logic*. George Boole, an English mathematician, introduced this algebra of logic in his book, *The Laws of Thought*, in 1854, thus the term *Boolean algebra*.

This algebra of logic can be used in the study of electrical circuits and switching. And through electrical circuits and switching,

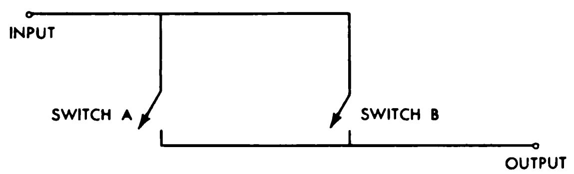


SWITCH A	SWITCH B	OUTPUT
0	1	0
0	0	0
1	0	0
1	1	1

Table of truths for series switching which simulates truth value of statements that use "and"

it is possible to study the algebra of logic. By setting up appropriate circuitry, the truth values of such statements as would imply connectives (or, and, not, etc) can be determined mathematically. Each of these connectives can be represented by circuitry, and the conditions of the circuitry designated by a number. For example, if the truth of a statement is shown as the digit "1" and the falsity of the statement as a "0," certain facts concerning the switching circuit to the left below can be determined.

In this circuit if a switch is open, it will be shown in the table as "0" (current is not



SWITCH A	SWITCH B	OUTPUT
0	1	1
0	0	0
1	0	1
1	1	1

Parallel switching circuit, showing truth value of statements that would use "or"

able to pass); if it is closed, it will be shown in the table as "1" (current may pass). In the output, the "0" and "1" will indicate a no-flow of current.

By referring to the table on the left, you find that there is just one condition of the switches that will allow current to flow to the output. This condition is when switch "A" and "B" are closed. Here you have simulated the truth value of statements that use the connective *and*. You find the digital notation to be "1 — 1 — 1."

To determine the truth value of statements that would use the connective *or*, a circuitry similar to the parallel switching circuit above can be set up.

In this case, as contrasted to the conditions of the series switching circuit, the current flows in all conditions of the switches except one — thus, the truth value of the statement "current will flow in the output if switch 'A' or 'B' is closed" has been determined.

The truth or falsity of these statements can be indicated in a digital manner by the use of a computing device that can scan the truth values of statements containing a variety of logical connectives.

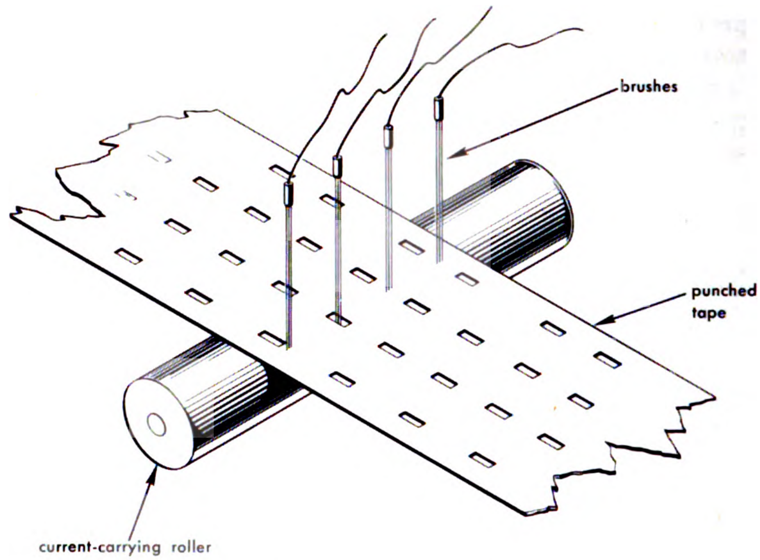
The results of a digital computation in a pilotless aircraft may be used in any phase of the flight. For example, data (statements) concerning the airspeed and altitude may be entered into a computing system to regulate the amount of fuel flow to the engine. Or perhaps the point is reached where the engine is to be shut off, and the missile is to dive into the target; the computer is called upon to determine the proper time for this to happen.

We have thus far covered the requirements and principles involved in having an electronic device that makes accurate computations in fractions of a second. Let's now take up the actual devices and some of the basic circuitry that make up a computing system. We will discuss these circuits from the overall standpoint of their possible uses, for it is next to impossible to list all of the applications of any one device or to say for sure that one particular device would satisfy the requirements of a problem better than some other device.

Input Unit

All the known data (statements) of a problem are injected into the computing system via the input unit. This input is usually supplied to the system by means of punched paper tape, photographic tape, magnetic tape, or wire on which the data and instructions have been recorded.

Tapes. In using a punched paper tape, such as illustrated on the next page, the original information is transferred to the tape according to a pattern that represents certain digits used in the computer. In the case of the binary digital computer, these digits would be "0" and "1" (indicating the various powers of two). The punched tape passes through a *reading unit* that permits electrical contact to be made whenever a hole is present under one of the current-carrying brushes.



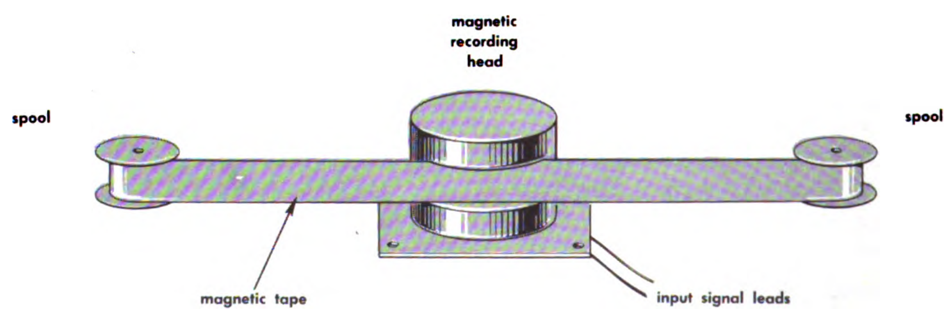
Punched-tape system of digital computers

The punch card or paper tape sorts, lists, selects, and copies information. It makes comparisons and selections according to instructions, and it adds, subtracts, multiplies, and divides. In order to use a punched tape, the original information is converted into patterns on the tape by means of some typewriter-like device. You may be familiar with the use of photographic tapes in the motion picture industry. In this case, the sound patterns are recorded on film, then photoelectric tubes are used to take this information from the film.

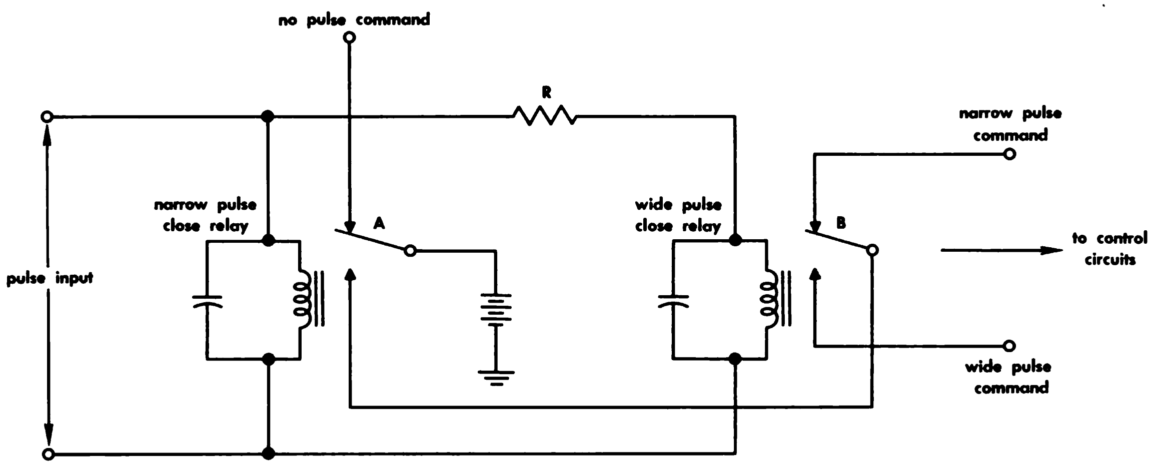
Another means of supplying the computer with information is by the use of magnetic tape or wire, upon which alternating currents have been recorded. In this case, the variations of the recorded signal bring about the operation of various components. The magnetic tape and wire *reading* (or pickup) units are quite similar to those of the commercial varieties. Note the figure below.

In addition to the recorded types of information that may be employed at the input, it is possible to have an operator originate the instructions to the computer. This could be accomplished by using push buttons, switches, or some other means of supplying data and instructions to the computer.

In missile equipment, the computer may receive data originating from error signals. These errors could be fed directly to the computer so that rapid corrections could be made.



Magnetic-tape recording



Pulse-decoding delay circuit

DECODING. As stated earlier, the process of decoding information is an important function that must be carried out by the computer. Information received at the computer must be put into a form that is suitable for the operation of other circuits. This preparation of information may be accomplished by any of several types of demodulation that were taken up earlier in the text. A type of decoding that lends itself well to a digital type computer is that of the *pulse-decoding relay* pictured on above.

The figure illustrates a method by which a command signal composed of narrow and wide pulses may be decoded in a relay circuit. When narrow pulses with long spacing are received, relay "A" closes. The second relay "B" does not close since the duration of the pulse is of insufficient time for its shunt capacitor to charge through the series resistor "R."

When signals consisting of wide pulses with short spacing are received, both relays close. The operation of these types of relay action are covered more thoroughly in the chapter dealing with relays.

Here, as in other digital devices, a means of forming a logical pattern is available. In this case, the connectives *if* and *and not* are introduced. This can be explained by considering the time when the pulses are applied. *If* a narrow pulse comes in, relay "A" closes *and not* relay "B."

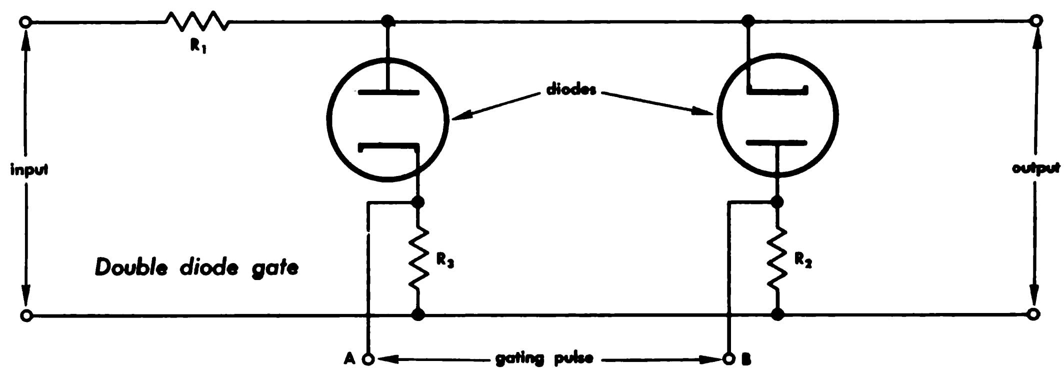
In the case of the wide pulse input, a connective of the form *both* is used. *If* a wide pulse is applied, *both* relays close.

In addition to the guidance and reference signals, signals may be fed back from other points of the guidance system. The computer input must, therefore, be capable of handling a large quantity of information.

Logical Unit

As you found out earlier, the simplest type of computer is one in which the basic components have but two positions. This simple computer made it easier to adapt to a binary numbering system rather than a scale-of-tens system. The switch was given as the basic example of a device that easily notes the condition of either *go* or *no go*. Other units, therefore, are designed to simulate the action of a switch. These components that exhibit two stable positions are generally referred to as *gates* or gating circuits. The gate, as the name implies, permits or prevents the flow of current in a circuit.

GATES. A fundamental method of electronic gating is by the use of a vacuum tube. Since a diode tube permits current to flow in only one direction, it may be used satisfactorily as a gating device in the computer. Current flows if appropriate voltages are present on the diode. This flow, in turn, causes other circuits to be influenced accordingly. Shown next page is a double-diode gating arrangement.



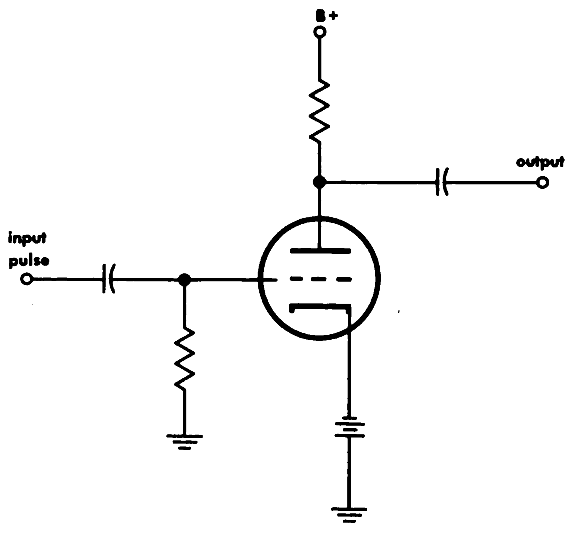
In a normal state, no current would flow in the resistances R_2 and R_3 and the output would be essentially shorted through these resistors and the diodes. When a current flows through R_2 and R_3 from point "A" to point "B," each diode plate is below its cathode potential, and the diode is effectively a large resistance. The output voltage in this arrangement will be a considerable fraction of the input voltage.

In addition to diodes, almost any type of multielement tube could be used as a gating device. A triode tube, biased at cutoff, would exhibit the characteristics desired of a gate, in that current would flow only during the times the grid was driven in a positive direction by the input signal.

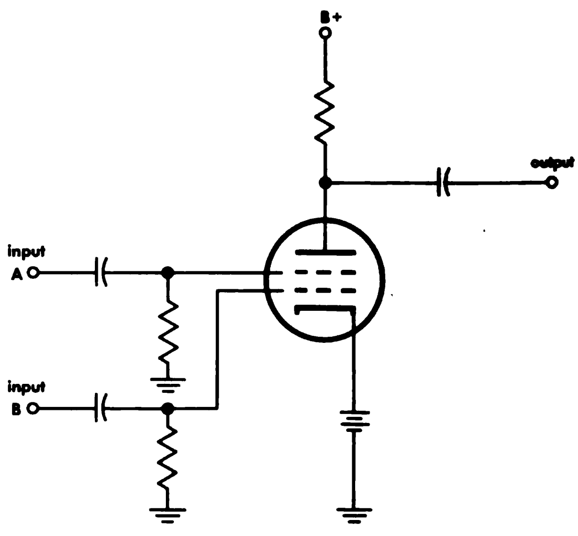
A tetrode could be used as a gating device. In such a device, the gating operation could be adjusted as desired. For example, the first (control) grid could be connected to the input, and the second (screen) grid could be connected to a bias voltage. A *control voltage* applied to the screen grid could bring the tube into a region where it may conduct. The input signal that was applied to the first grid could then make itself apparent at the plate (output) of the tube. The triode and tetrode gating systems are illustrated below.

Other common types of gates employing vacuum tubes are the parallel gate, the common-cathode gate, and the diode-coupled grid gate, the circuitries of which are shown on the right.

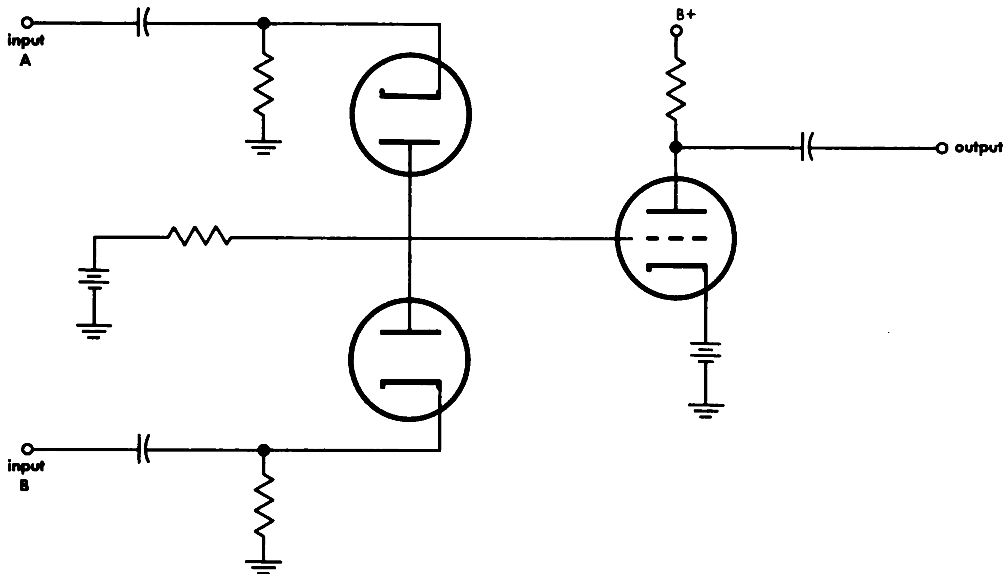
COUNTING. Another important function that takes place in the logical unit is that of *counting*. In any computation, it is necessary



Triode gate



Tetrode gate



Diode-coupled gate

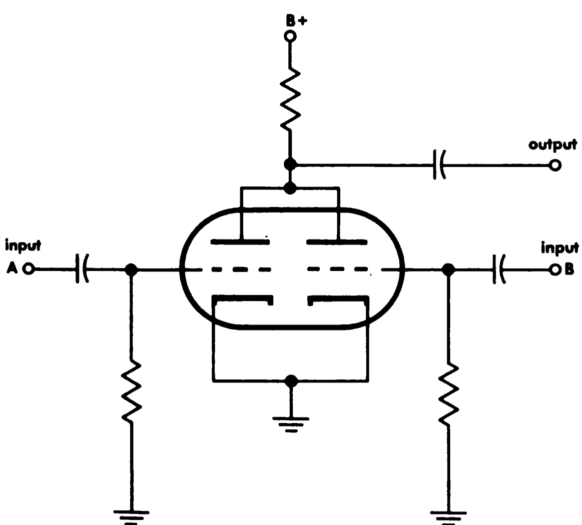
to have a means by which the number of operations may be determined. This is the function of the *counter*.

This function is performed in connection with the punched tape or the magnetic recording. In these cases the repetition of pulses would be recorded along with the other operational functions.

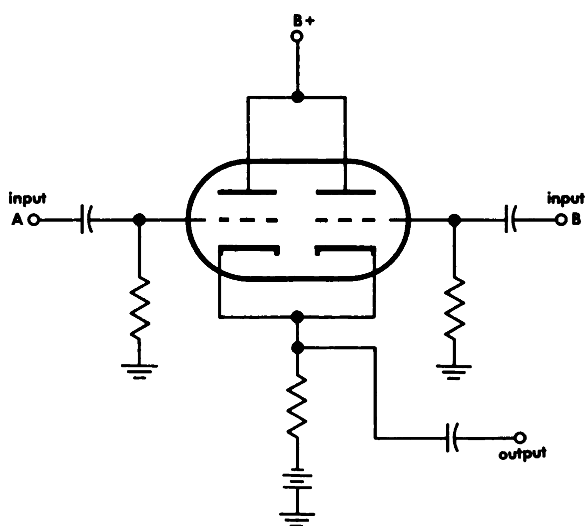
Once again, a electromechanical switching principle may be utilized in the functions of counting. In this case, a multiposition switch is actuated by some electromechanically operated clutch system to give any desired number of pulses in a given period of time.

The counting function may or may not be a continuous operation in a computing system. When counting takes place depends on the state in which other components of the system are set — that is, it may be necessary to count a certain number of other operations that have taken place before some new operation is to begin.

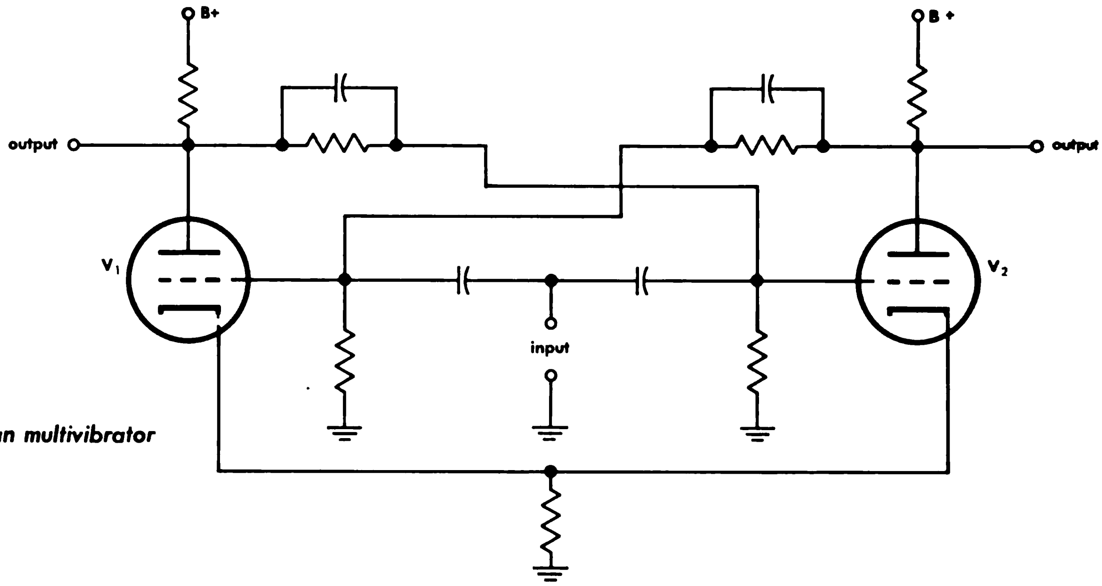
One of the basic means of having either a continuous or interrupted system of counting is by employing the multivibrator or flip-flop principle. The most common form of the multivibrator uses two triode tubes, having plate-to-grid feedback from one to the other



Parallel gate



Common cathode gate



Eccles-Jordan multivibrator

and a common bias arrangement, between the two halves of the circuit. The Eccles-Jordan multivibrator, shown above, illustrates the flip-flop principle. This multivibrator may assume either of two stable conditions.

Referring to the figure of the Eccles-Jordan multivibrator, one of the stable conditions it may have is when tube V_1 is conducting and V_2 is cut off; the other stable condition is when V_2 is conducting and V_1 is cut off. Let's consider the condition when V_1 is conducting and V_2 is cut off.

Suppose that during the assumed condition a negative signal appears at the input terminals. This negative signal appearing at the grid of V_1 causes a decrease in the current flow through this tube. This decrease in current flow permits the plate voltage of V_1 to increase in a positive direction. The positive increase of plate voltage at V_1 is reflected on the grid of V_2. As the grid of V_2 is driven in a positive direction, the tube begins to conduct, and its plate voltage is reduced (or driven in a negative direction). The negative going voltage is reflected upon the grid of V_1, reducing the conduction of this tube. As the V_1 plate voltage is increasing, the V_2 plate voltage is decreasing until a point is reached at which the feedback of voltage causes V_1 to be cut off and V_2 to be conducting. This process would be reversed at the time when another negative input pulse is introduced to the circuit.

You can see how this multivibrator circuit lends itself well to a binary system using the digits 0 and 1. Nonconduction of one of the tubes could represent the "0," while the conduction state could represent the "1."

There are many adaptations of the flip-flop principle; and, as you will see in the next part, this principle can be used equally as well in the memory circuits of the computer.

Memory Unit

A memory unit stores information until it is called for in the computation. A memory unit is actually the "backbone" of the computer. The term *delay circuit* is often used in dealing with the memory unit of the computer. However, this delay is in actuality a form of storage — a short-time storage system. The usage of the term *delay* is usually with regard to a particular operation that is to be performed after some other operation. Storage, on the other hand, refers specifically to the retention of some information that will be needed in later calculations.

Here again, in memory circuits, we discover the diversity of components used in a computer. Some of the components that were common to the input and logical units are also used in the memory circuit. This is true of the punched tape, magnetic recording, and the switch. There are some circuits that are quite limited as to their uses; we will consider some of these that find their specific use in the memory unit.

One of the most adaptable means of high-speed reference storage is the sonic delay line illustrated below. Sonic delay lines, as the word implies, are lines that function by transmitting sound pulses through a medium of liquid, solid, or gas. The one in the illustration is a liquid-type delay line, using mercury as the transfer medium.

In the above delay line, two quartz crystals are used, one as a transmitting element, the other as a receiving element. Quartz crystal is an efficient material in that it exhibits excellent piezoelectric effects, and the acoustical impedance of quartz and mercury are comparable. Matching of acoustical impedances indicates that there will be an optimum transfer of sound energy relayed between the mercury and quartz. Suppose, for example, that a signal is fed into the delay line and that it has the coding of 101011 (in accordance with the binary system). The quartz crystal receives this code as "pulse, no pulse, pulse, no pulse, pulse, pulse." The quartz crystal in turn sets up sound waves within the mercury that have the same rhythm, or repetition. The sound waves in striking the quartz crystal at the opposite end of the delay line reproduces the electrical impulses. These pulses of energy are then amplified and returned to the input of the delay line where this cycle of energy transfer is repeated.

This cycle of transmitting and receiving the impulses is continued until such time that a gating circuit is operated that will either transfer the energy to other components or clear it from the system.

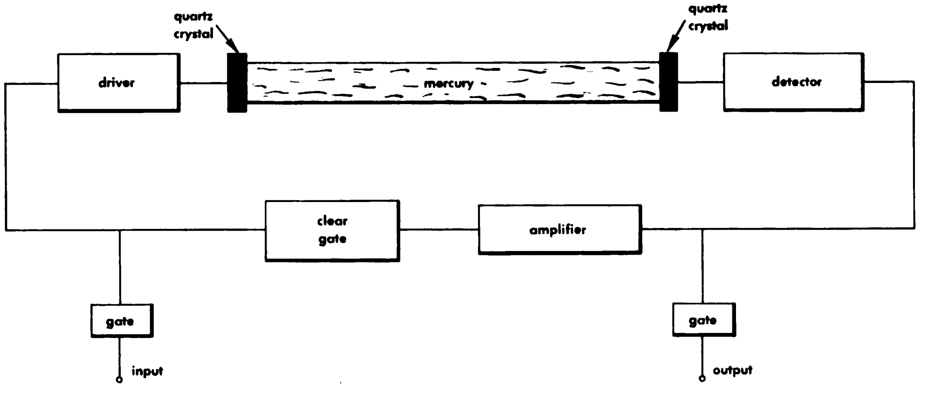
Another means of storing information is by electrostatic storage — cathode-ray-tube storage. Shown on the next page is the basic electrostatic storage tube. In this case, a beam of electrons is directed upon a surface composed of conducting and nonconducting layers. As the electrons strike this surface, an electrical charge is created. The presence or absence of the electrical charge gives an indication of 0 and 1 conditions. Thousands of units of information can be handled readily by the electrostatic tube. The information can be stored or erased readily, and in conjunction with other circuits this information can be retained for long periods of time.

Phosphorescent materials are also used for digital storage. In this case, a cylinder (drum) is coated with a phosphorescent material and energized with electromagnetic radiation. The irradiating spots are then read off by means of a photoelectric cell. Due to difficulties with light shielding, heat interference from the system, and optical interference, this system of storage is limited in application.

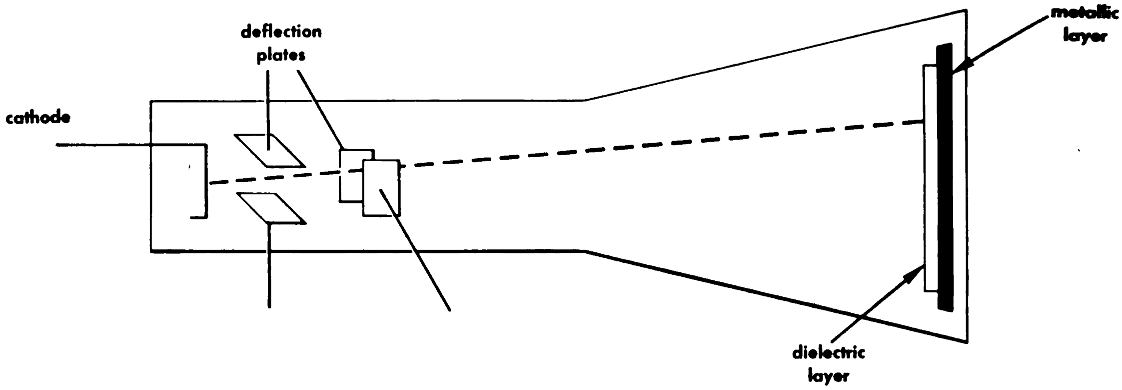
Output Unit

After all computations have been completed, there must be some means of presenting the answers to the operator. Any of the devices that were used for putting information into the computer may also be used to take information from the machine. This includes recorded tapes of all types and relays, whose positions may be indicated by light bulbs.

Geared indicators, similar to those on the common desk-type calculator may also register the results.



Mercury-tank sonic delay line



Electrostatic storage

In an airborne missile system, the results of the computer may be used directly to influence the control system or perhaps the sensor unit in a directional system of guidance.

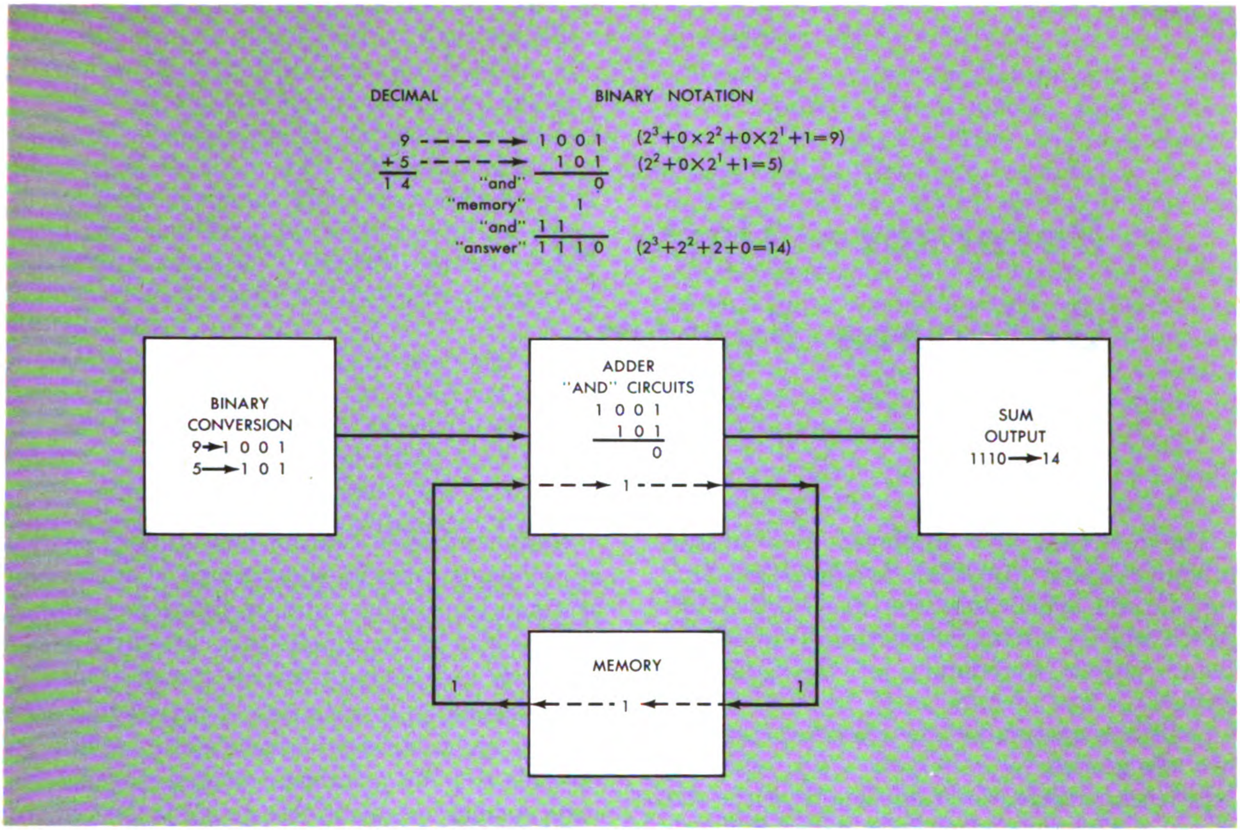
Mechanical Systems in Digital Computers

At this point we have taken up the operational techniques and a few of the common types of components used in the digital computers. The following table lists the various functions of a digital computer along with the common types of devices that may be utilized in each.

It should be noted that during the discussion no mention was made of any mechanical systems used within the digital computing machines. These were not omitted because of unimportance, but because of the fact that the mechanical types are generally peculiar in their construction to fit a particular job. For example, it may be desirable to use some type of a gearing mechanism to perform a calculation. However, this gearing may involve several types and arrangements to fit the purpose.

Many mathematical functions such as addition, subtraction, multiplication, or division may be handled within the computer. Consider the addition processes performed by a computer. The first step in making ready the information for the binary digital computer is to put it into the binary numbering system. This binary form of the information is then fed into circuits that perform the logic operation of *and*. The memory units come into operation whenever there is any *carrying* to

INPUT UNIT	LOGICAL UNIT	MEMORY UNIT	OUTPUT UNIT
Tapes	Switches	Tapes	Tapes
Punched paper (or cards)	Gates	Punched	Punched
Photographic	Vacuum tubes	Magnetic	Magnetic
Magnetic tape or wire	Tapes	Photographic	Photographic
Buttons	Punched	Sonic delay line	Switches (lights)
Switches	Magnetic	Electrostatic	Relays (lights)
Relays	Photographic	Phosphorescent	Geared indicators
Mechanical	Flip-Flop (multivibrator)	Mechanical	Mechanical
Error Signals from missile	Relays		Missile correction of errors
	Mechanical		



Binary-digital addition

be done. For example, if the two decimal numbers, 9 and 5 are to be added, they are converted to their binary equivalents 1001 and 101 respectively. Then the last two digits of the binary number enter an *and* circuit. This 1 plus 1 gives the number 2 or, in binary notation, 10. Therefore a "1" has to be stored for a short time. Further *and* circuits then operate to give a final binary answer of 1110. This process is illustrated above for further clarity.

The process of subtraction would be similar to that of addition except that circuits employing *not* and *or* logical patterns would have to be utilized. These circuits may be using flip-flop circuits or relay systems in addition to adders. Another method of subtraction in computers is to take complements of numbers, and then perform algebraic-addition operations.

Any multiplication process can be, and usually is, performed by a series of additions. Likewise, the division of a group of numbers may be accomplished by the repeated process of subtraction.

Any mathematical process can be duplicated by the use of digital computers. Problems involving square roots, trigonometric functions, or differential equations can be handled with great speed and accuracy.

COMPUTING DEVICES IN THE MISSILE FIELD

In the missile field, many computing devices are at work. Some of these are quite obvious and are installed as separate units, while other computing components may not be readily apparent. Future study of computing systems will be easier for you if you keep in mind that all missile electronic systems utilize computing devices; computing devices are there whether they are separate units or hidden components.

Reference Units of Guidance Systems

The reference unit employed in a missile guidance system provides standards for generation and synchronization of electrical impulses and for timing of electronic circuits to insure proper functioning of the guidance system.

Guidance references are classified broadly into two groups:

External references (associated with the base station, ground radar, and transmitters)

Internal references (located within the missile)

EXTERNAL GUIDANCE REFERENCE UNITS

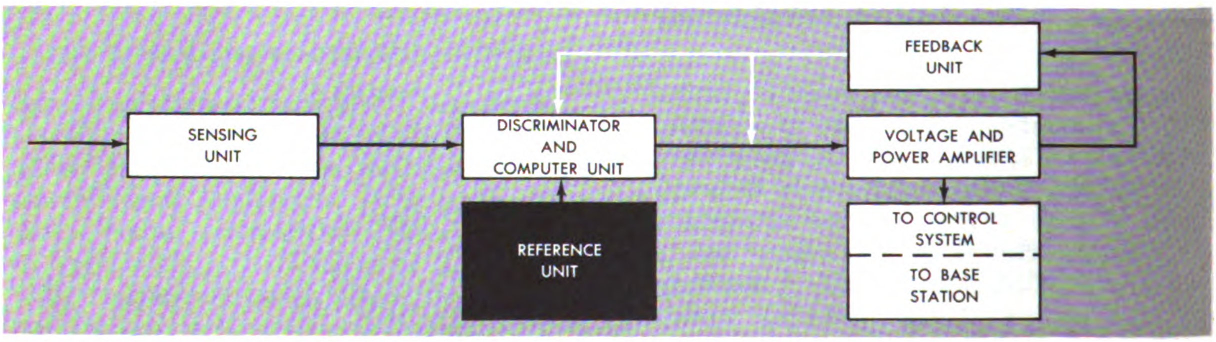
External guidance reference units are considered to be those that are outside the missile but essential to the operation of the system in controlling the course of the missile. Included in this category are radio and radar beacons, celestial bodies, and topographic or gravitational characteristics of the earth.

Radio and radar beacons find primary usage as reference units in short range missiles. In systems employing radio or radar reception, the position of the missile is accurately established by one or more beacon transmitters.

Reference, in some cases, is established by *homing* on an enemy transmitter. It also is possible to *plant* a transmitter in enemy territory. A small missile is used to plant the transmitter. After the transmitter is planted, a large missile carrying a heavy payload can be accurately guided to the target.

Hyperbolic grids (navigational networks) are another type of radar external references. These references, however, are generally used in establishing the initial part of the flight only.

Celestial bodies also afford an accurate external reference. By means of precise star tracking systems, along with predetermined



Basic diagram of missile guidance system

information, a missile is programmed throughout its course. Stars afford an excellent reference for long-range missiles.

The topographic and gravitational characteristics of the earth are additional reference standards that are used either for short- or long-range systems. Topographic features of the earth are used for reference in all map-matching systems. Television systems may also be considered as utilizing the topography of the earth.

Gravitational influences of the earth serve to establish the reference for all of the inertia types of guidance systems. Here, plumb bobs, gyros, accelerometers, and other inertial devices establish the position of the missile by making reference to the gravitational influence of the earth.

The various units that may be used as external reference shall not be taken up in detail here. Some of these units are gone into in a later discussion on guidance systems. The rest of this discussion of reference units relates to the internal reference units — those found within the missile itself.

INTERNAL GUIDANCE REFERENCE UNITS

Although some of the internal reference units considered here, such as stabilized platforms, make use of outside references, they are covered as internal reference units because they are *within* the missile.

Timing Controls

Let's first consider the various methods of timing. One of these methods or various combinations of these methods might be used in a guidance system.

OSCILLATORS. Oscillators used as timing controls and frequency standards must possess a high degree of stability under all operating conditions. This means that they must be relatively insensitive to variations in power-supply voltages, temperature, and pressure.

They must provide output voltages of constant frequency and amplitude to serve as standards or time bases for such guidance-system units as trajectory playbacks, scanner motors, fixed fields of gimbal and platform torquers, gyro rotors, gyro torquers, and microsyn signal generators. All of these applications demand a high degree of accuracy at relatively low operating frequencies.

Crystal-controlled oscillators possess most of the desired features except that of low-frequency operation. Due to piezoelectric characteristics, they are restricted generally to frequencies of 100 kc or higher and are susceptible to variations in temperature.

Relaxation-type R-C oscillators employing gaseous tubes operate at very low frequencies but tend to be somewhat erratic. They require a constant-voltage power supply. Multi-vibrator-type oscillators also are subject to these latter limitations.

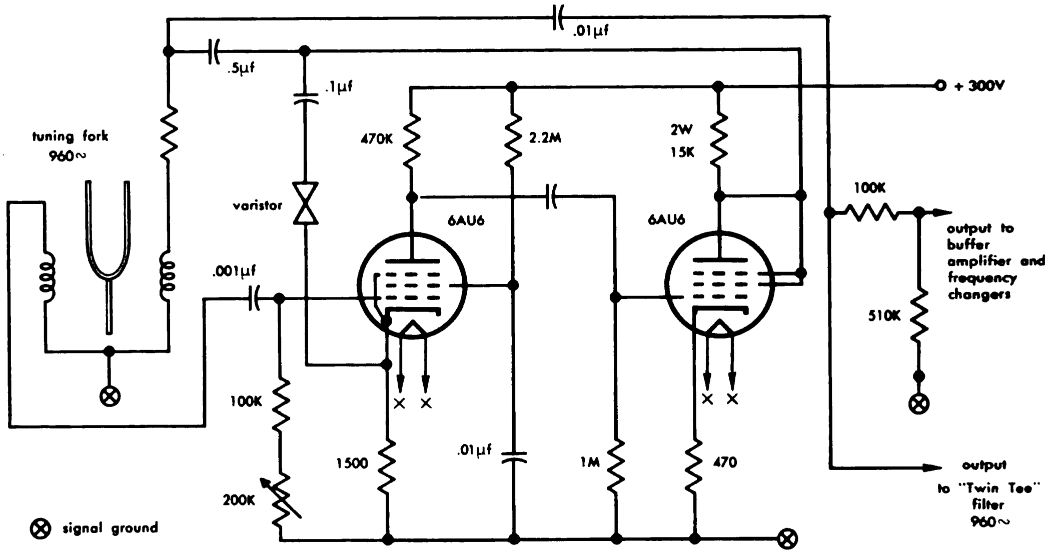
To meet the requirements of a missile guidance system reference, many modifications of the foregoing oscillators have been tried. Two such modifications are crystal-controlled oscillators with frequency dividers and multi-vibrators with regulated and compensated power supplies.

One of the most satisfactory time-base generators developed to date is a *tuning-fork oscillator*. At a frequency of 960 cycles per second, this type of oscillator can be stabilized to a small frequency error over a wide range of temperature.

The tuning-fork oscillator diagrammed on the next page is controlled by a 960-cycle tuning fork in the grid and plate circuits of the 6AU6 oscillator. As the tuning fork vibrates, the inductance of the coil in the grid circuit changes, producing a signal on the grid of the 6AU6 pentode. This signal is taken from the plate of the pentode and coupled to the grid of the triode-connected 6AU6 amplifier.

From the plate of the triode-connected 6AU6 amplifier, the signal is capacitively coupled back to the driving coil of the tuning fork which is wound to produce a voltage of the same phase in the grid winding. This provides the regenerative feedback necessary to sustain oscillation.

A portion of the signal on the plate of the triode-connected 6AU6 is coupled through a 0.1-uf condenser and a varistor, then back to the cathode of the pentode 6AU6. This provides degenerative feedback to control the oscillator. If the oscillations at the plate of the triode tend to increase due to a change in temperature, or some other transient condi-



Tuning-fork oscillator

tion, more degenerative voltage is fed back to the cathode of the pentode stage, thus decreasing its gain.

The 200-kc potentiometer in the grid circuit of the oscillator controls the amplitude and phase of the grid voltage. This serves as a frequencv adjustment, preventing a frequency change that may be caused by too much current in the driving coil of the tuning fork.

One sinusoidal output from the 960-cycle tuning-fork oscillator is fed through a *twin-tee* or *parallel-T* filter designed to pass only a 960-cycle signal. The output is then applied to a gyro signal generator or other circuitry which utilizes synchronizing or timing signals of this specific frequency.

Another output is passed through a buffer amplifier to frequency-dividing or frequency-multiplying circuits to produce other timing frequencies such as 60, 120, 240, 480, 1920, 3840 cps, etc.

Various other time-base generators suitable for applications requiring a lesser degree of accuracy are discussed later in the text.

FREQUENCY CHANGERS. Frequency changers are either mechanical or electronic devices used to divide or multiply the output frequency of a standard signal generator. If the application requires a signal frequency within close tolerances, it is essential that the frequency-changing device possess the same degree of accuracy as the standard time-base generator. This requirement restricts the use of mechanical devices such as choppers and vibrators as timing controls or frequency standards in missile guidance reference units.

Where even submultiples are desired such as f/2, f/4, f/8, etc, of a standard reference frequency, a bistable multivibrator can be used as a frequency divider. This device produces a square-wave output with a frequency equal to one-half that of the input signal, and either half of the multivibrator may be conducting when no input signal is present. Two or more such devices can be cascaded to further produce even submultiples of the standard frequency signal. A bistable multivibrator for dividing frequency by two is shown on the right.

"R-1" and "R-4" provide a DC coupling between the grid of "V-1" and the plate of "V-2." "R-2" and "R-3" provide a DC coupling between the grid of "V-2" and the plate of "V-1." If "V-2" is conducting, its lowered plate voltage is applied to the grid of "V-1" through "R-4" and "R-1," producing a net voltage of 15 v to -20 v on the grid of "V-1."

A negative 960-cycle input signal causes "V-2" to cease conducting and causes "V-1" to conduct. Another negative input signal is

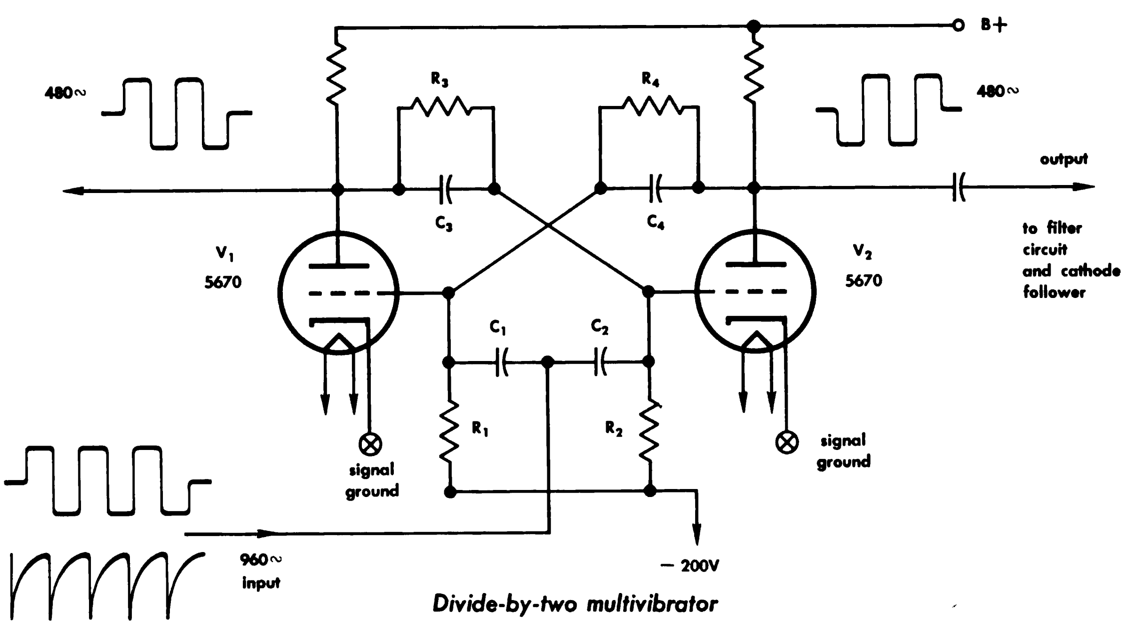
required to "flip" the circuit again by causing "V-1" to cease conducting and "V-2" to conduct again.

Thus, the action of this circuit divides the 960-cycle input signal into 480-cycle square-wave signals at the plates of "V-1" and "V-2." These signals are coupled to the grids of cathode-followers through harmonic filters so that 480-cycle sine waves are obtained from the cathode followers for reference purposes in synchronous and test goniometers, guidance pulse generators, or any other circuitry requiring a reference signal of 480 cycles per second. This 480-cycle signal can be applied to a similar circuit for producing a 240-cycle reference signal, and so on down to frequencies of 60 or 30 cps, below which this circuit tends to become too unstable for reference purposes.

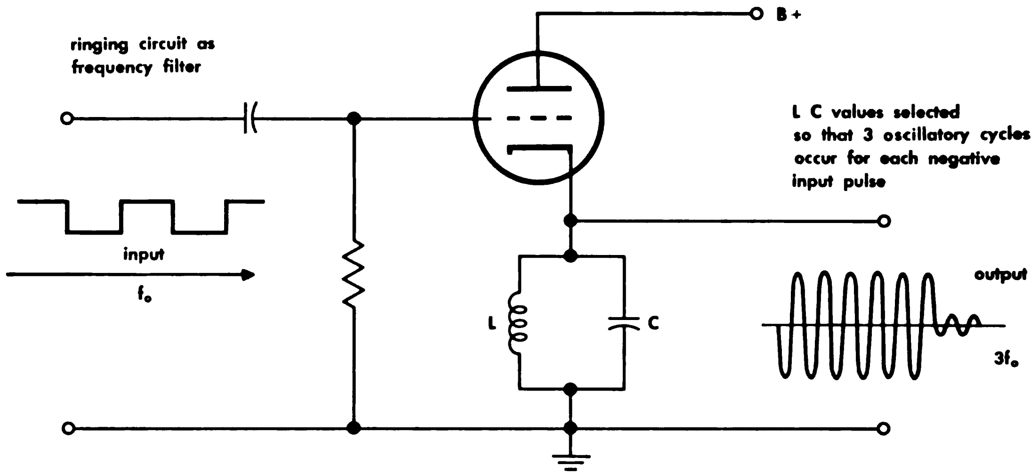
Frequency multipliers also may be mechanical or electronic; but, as in the case of frequency dividers, they must possess stability and accuracy in the same degree as the primary frequency reference when used as standards of reference. Few mechanical devices possess these desired qualities and, consequently, are not suitable for application in missile guidance reference units. At radio frequencies, electronic multiplication is relatively simple. It is accomplished by the use of amplifier stages with tuned grid and tuned plate circuits. The grid circuit energized by the primary reference signal, and the plate circuit is tuned to a harmonic of the fundamental frequency. This system is widely used. Where power requirements are small, harmonics as high as the 55th have been used. However, any inaccuracy in the primary signal is multiplied proportionately in this type of frequency multiplier and may render the signal unsuitable as a reference.

One method of frequency multiplication which is widely used for producing time-bases and range markers in radar equipment employs a circuit known as a *ringing circuit*, or ringing oscillator. The ringing circuit employs a suitable amplifier tube with a tunable LC circuit in series with its cathode. The tube is normally conducting until the negative portion of the signal cuts it off. At this time, the field produced by the normal cathode current, flowing through the inductor (L), collapses. Its current flows into the capacitor (C), initiating an oscillatory action. The values of "L" and "C" determine the frequency of oscillation; therefore, the primary reference signal is used to start and stop the ringing action of the LC tank circuit.

The number of oscillations produced in the tank circuit is controlled so that the frequency of the oscillations is any desired multiple of the frequency of the triggering signal (primary reference). The controlled frequency must be within the tuning range of the LC tank circuit. A typical ringing circuit is illustrated on the following page.



Divide-by-two multivibrator



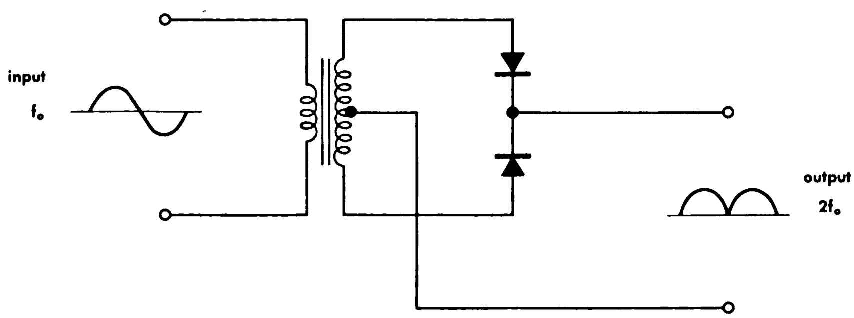
Ringing circuit

A simple, efficient frequency doubler which is suitable for use with a primary signal source having a sinusoidal output is the familiar full-wave rectifier circuit, shown below.

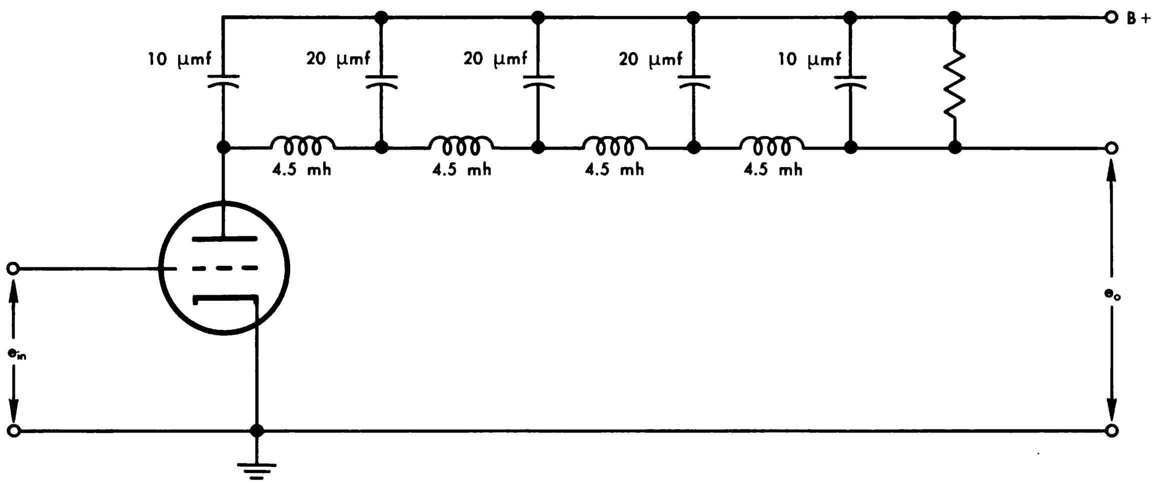
DELAY CIRCUITS. Missile guidance systems frequently employ circuits which require a finite time for their actuation. Therefore, the timing and sequence of operation of each circuit must be closely controlled with respect to the master signal or data pulse. To obtain the desired sequence of operation, the pulse or wave-form is delayed or stored by means of some mechanical or electronic device which has a definite period of transmission between its input and output. An electronic circuit which performs this function is called a *delay* circuit or *storage* network. Delay circuits also are employed in radar applications to measure time intervals and form pulses, to synchronize sweep circuits and time-base calibration in oscilloscopes and sweep-generators, and to provide channel separation in time sequence in multiplex communication systems.

The type of delay circuit employed depends on the characteristics of the data to be delayed and the time delay required. Some circuits commonly used as delay circuits are artificial transmission lines, *start-stop* multivibrators, and the phantastron circuit. These circuits and their applications are discussed in this chapter.

Since any conductor possesses inductance and capacitance, a transmission line is considered to be an LC circuit. The characteristics of a transmission line can be duplicated by proper combinations of "L" and "C" used as a delay circuit suitable for delay time within the range of 0.01 to 100 microseconds. Some artificial transmission lines consist of wave-



Full-wave rectifier as frequency doubler



Circuit employing an artificial transmission line

filter sections mutually coupled. For long time-delay requirements, however, so many sections might be needed to provide the required delay and sufficiently high cutoff frequency, that some other delay device might be more practical.

For delay time longer than 100 microseconds and up to several thousand microseconds, a pulse is delayed by transmitting it via supersonic waves through a liquid medium, such as mercury. In some applications, such as electronic computers, pulses are stored in mercury *ultrasonic tanks* for relatively long periods of time.

Transmission lines, such as coaxial cables and parallel conductor lines, are used as delay lines when delays of the order of a fraction of a microsecond are required.

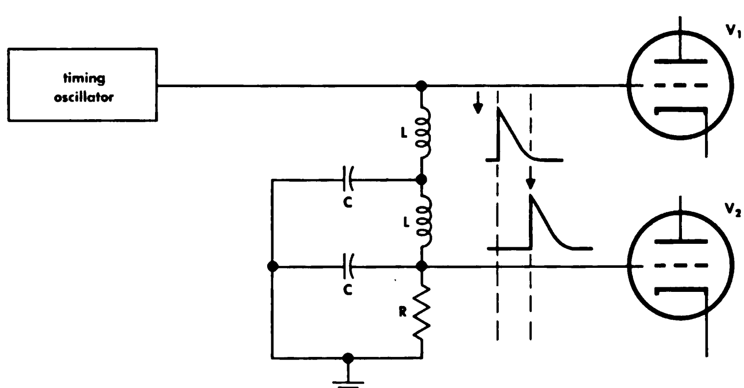
The time delay (Td) of the line above can be calculated by means of the equation:

$$Td = N\sqrt{LC}$$

where "Td" is the time delay in seconds, "N" the number of sections, "L" the inductance per section in henries, and "C" the capacitance per section in farads. "Td" is solved for as follows:

$$Td = N\sqrt{LC}$$
$$Td = 4\sqrt{4.5\times10^{-3}\times20\times10^{-12}}$$
$$Td = 4\sqrt{9\times10^{-14}}$$
$$Td = 4\times3\times10^{-7}$$
$$Td = 12\times10^{-6} = 1.2\times10^{-6}\text{ sec}$$
$$Td = 1.2\text{ microseconds}$$

The characteristic impedance "Zo" of the delay line in the above example is found from the formula:



Artificial transmission line in which V_2 operates at later time than V_1

$$Zo = \sqrt{\frac{L}{C}}$$

where: L = inductance per section in henries

C = capacitance per section in farads

$$Zo = \sqrt{\frac{4.5 \times 10^{-3}}{20 \times 10^{-12}}}$$

$$Zo = \sqrt{\frac{4.5}{2 \times 10^{-8}}} = \sqrt{2.25 \times 10^{8}}$$

$$= 1.5 \times 10^{4} = 15{,}000$$

$\therefore$Zo of the delay line = 15,000 ohms.

If this delay line is terminated in a 15,000-ohm resistor, pulses applied to the input of the line by the tube travel to the output end, and are not reflected. The operation of the tube is the same as if a 15,000-ohm resistor were connected directly in its plate circuit.

When an artificial transmission line is terminated in its characteristic impedance, voltages applied at its input end are dissipated by the resistance termination without reflection. If the line is open-circuited at its terminal end, any voltage applied to the input is reflected back from the open end without change of phase, just as in the real line.

The figure on the preceding page illustrates an application of an artificial transmission line used to introduce a delay so that "V_2" is operated at a slightly later time than "V_1." The delay is determined by the values of the lumped "L" and "C" constants using the previously mentioned formula:

$$Td = N\sqrt{LC}$$

where "N" is the number of sections of line used. Since "R" is the terminating resistance of the line as well as the grid input resistance of "V_2," it must be equal to the characteristic impedance of the line to prevent reflections.

An artificial transmission line has characteristics similar to low-pass filter. Such a filter tends to pass all frequencies below its cutoff value (Fc) and to reject all frequencies higher than the cutoff value. The greater the number of sections in the line, the more sharply defined is the cutoff frequency, which may be expressed as:

$$Fc = \frac{1}{\pi\sqrt{LC}}$$

Synchronizing pulses usually are either rectangular in shape, or they consist of a sudden rise in voltage followed by a slower exponential delay. Such wave forms include a considerable number of high-frequency components; therefore, if the delay line is to pass such a wave without distortion, its cutoff frequency must be high enough to pass the necessary frequency spectrum.

It has been proven mathematically that the time of rise for a pulse being passed through a delay line is limited to the reciprocal of twice the cutoff frequency, or:

$$\frac{\pi\sqrt{LC}}{2}$$

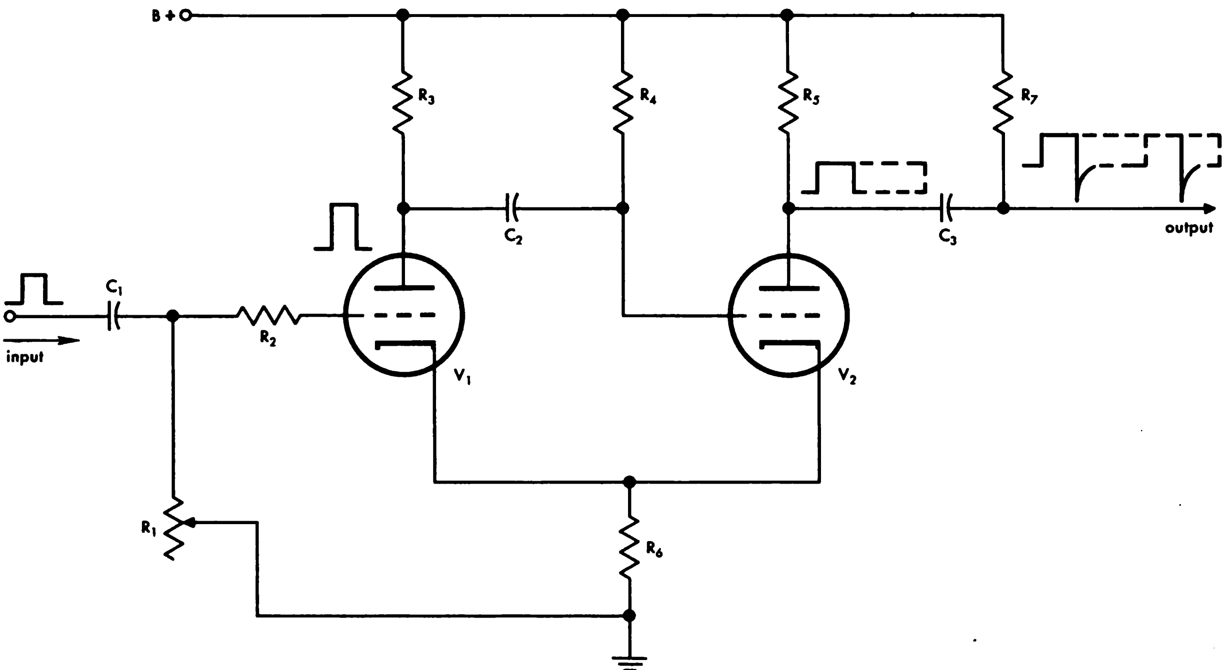
In the guidance systems of missiles, radar pulses are frequently used to initiate the operation of various circuits within the missile.

When the same pulse is used to trigger more than one circuit, and where some specific time interval or sequence of operation is desired between activation of the circuits, a variable delay circuit must be provided. The start-stop (monostable) multivibrator is suitable for producing the desired delay in many of these applications.

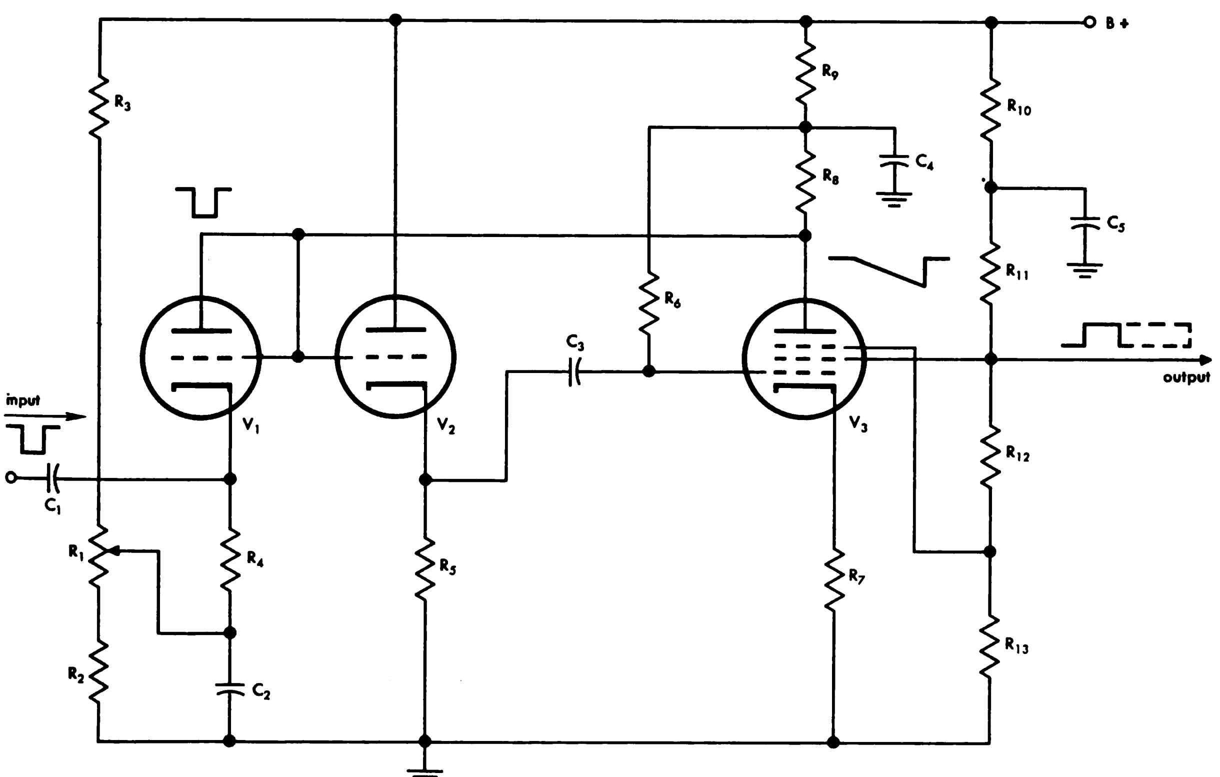
A monostable multivibrator capable or producing a positive square wave and a delayed negative pulse is shown in the accompanying figure. In this circuit, the components are so selected that with no input signal, "V_1" is cut off and "V_2" conducts.

When the positive signal pulse appears at the grid of "V_1," "V_1" conducts and "V_2" is driven toward cutoff.

The setting of potentiometer "R_1" regulates the discharge time of "C_1" and the time that "V_1" conducts. When "V_2" is driven to cutoff, its plate voltage rises. Due to the time constant of R_4" and "C_2", which form a Miller integrator circuit, a positive square wave appears at the plate of "V_2," the duration of which is determined by the RC network values and the setting of "R_1." This square-wave output is differentiated through a very short time-constant circuit consisting of "R7" and "C3," producing the negative peak or pulse which may be used to trigger a succeeding stage. The position of this negative pulse depends on the width, or duration, of the positive square wave at the plate of "V2." There-



Monostable multivibrator pulse relay circuit



Phantastron pulse delay circuit

fore, its occurence with respect to the input pulse is delayed by the duration of the square wave as controlled by "R_1."

The action of a Miller integrator is discussed in connection with amplifier units in the next section of this chapter.

PHANTASTRON CIRCUIT. The phantastron circuit is a medium precision delay circuit which is similar in operation and output to the "flip-flop" circuit. It possesses an advantage over multivibrator circuits in that it is quite stable under power-supply voltage variations.

This circuit is frequently employed as a delay circuit for timing the sequence of operations in missile guidance systems and in pulse-decoding systems.

The previous schematic illustrates a phantastron circuit employed to delay a negative input pulse over a range of approximately 50 to 350 microseconds.

Control-grid initiation of delay is used in this circuit instead of the conventional suppressor-grid method. Plate voltage is set by "R_1." "V_1" is conducting and holds the control of grid of the phantastron (V3) at a constant potential until the signal-pulse appears.

The suppressor-grid voltage is determined by the bleeder chain "R10," "R11," "R12," and "R13."

The time delay is adjustable by the setting of "R_1," which determines the plate voltage and the initial charge on "C3." "C3" with "R6" comprises a Miller integrator between the plate and grid. "V_1" serves as an input control diode and "V_2" as a cathode follower input.

The operation of a similar circuit employing the Miller integrator is covered in Section IV of this chapter. The Miller integrator circuit is used as a pulse shaper rather than as a delay circuit in the amplifying units.

GONIOMETER. A goniometer unit consists essentially of autosyn phase shifters connected by common gearing with external hand control for driving the gear train.

The signal input to the unit is 100-kc, 10-kc, and 2-kc sinewave voltages from the timing unit, while the output consists of the same sinewave voltages displaced in phase with relation to the input. This electrical phase shift is accomplished by means of an autosyn resolver for each frequency of the input voltage. The autosyns are connected by a gear train in such a manner that the electrical phase shift of the 100-kc sinewave voltage is ten times the phase shift of the 10-kc sinewave voltage and fifty times the electrical phase shift of the 2-kc sinewave voltage. The phase-shifted output voltages of the unit, when used in conjunction with the pulse generator, originates a video pulse which can be delayed in time by any amount.

A handwheel drive for the goniometer gear train is located on the front panel of the unit to provide a means for varying time delay. A second control is provided for step delays of 100 microseconds.

The 2-kc sinewave signal is fed into an impedance coil ("L_1"-"L_2" in the accompanying circuit diagram), which makes the circuit look like pure resistance, and into an RC network across the two coils of the rotor which are wound mechanically 90 degrees apart. The current through capacitor "C_1," the coil "L_1" of the autosyn, and the resistor "R_1" will lead the current by 90 degrees through resistor "R_2," capacitor "C_2;" and the other coil of the autosyn "L_2."

To prove that the voltages are 90 degrees out of phase: The capacitive reactance is 796 ohms; therefore, the currents in loop "A" of the diagram top right are equal to those in "B" displaced 90 degrees as follows:

Dividing: $\frac{I_1Z}{I_2Z_2} = \frac{Io\ (-JXC_1)}{IoR_2} = \frac{JXC_1}{R_2}$

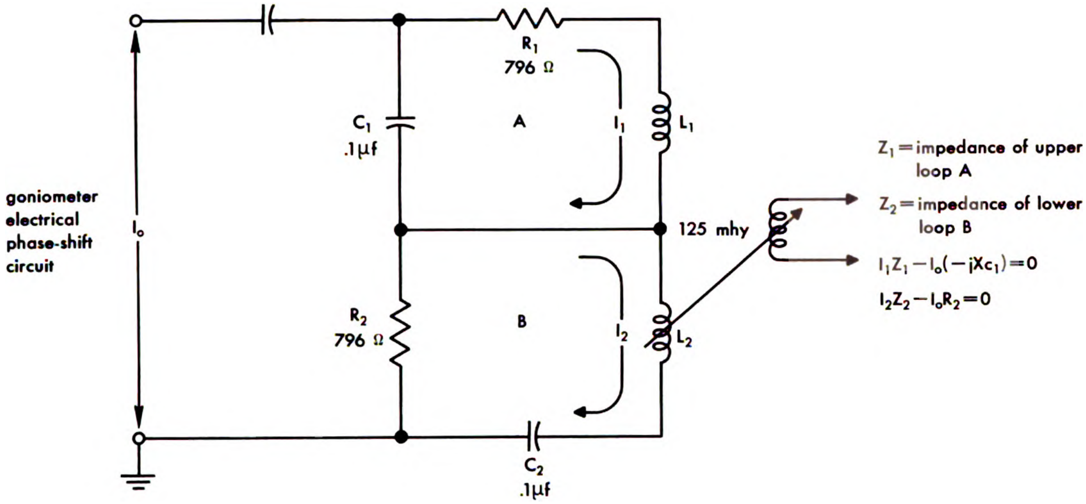
Since: $R_1 = R_2 = XC_1 = XC_2$; $XL_1 = XL_2$; $Z_1 = Z_2$.

Then: $\frac{I_1Z_1}{I_2Z_2} = \frac{JXC_1}{R}$; $I_1 = JI_2$; $EL_1 = JE_{12}$

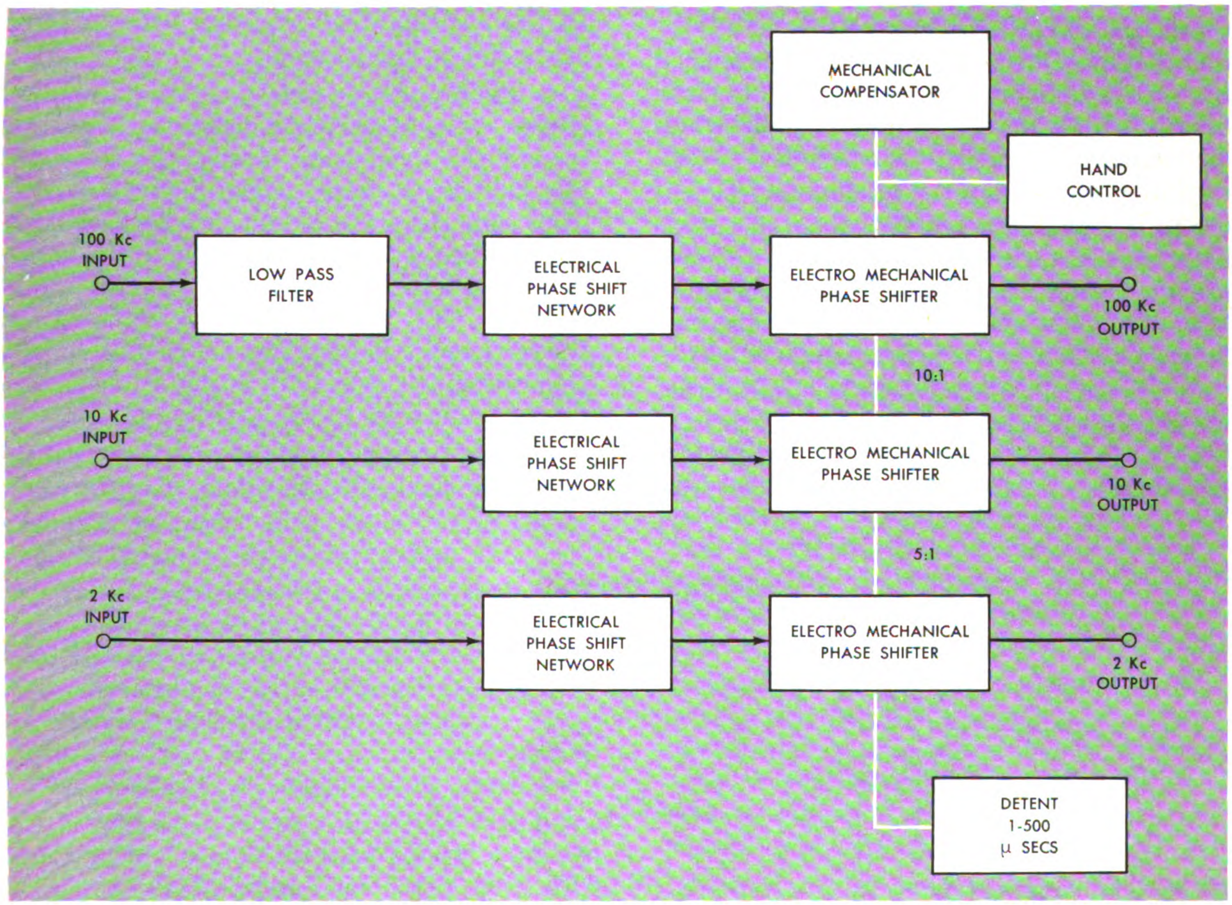
The above formula is the same for any of the three autosyn networks. The output voltages of constant amplitude are furnished to the synchronous pulse generator and the monitor scope.

The figure at the right illustrates in block-diagram form the goniometer unit discussed in the preceding paragraphs.

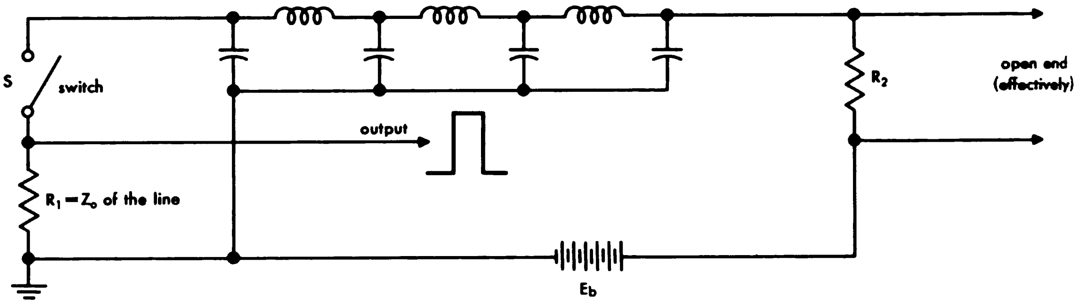
PULSE-FORMING CIRCUITS. In order to transmit short, powerful bursts of RF energy, the transmitting tube of a microwave radar sys-



Goniometer circuit



Goniometer unit diagram



Pulse-forming line with a mechanical switch

tem is modulated (or keyed) by high-voltage rectangular pulses of short duration. A pulse of rectangular shape is desirable because it allows the transmitter tube to operate with maximum efficiency and frequency stability.

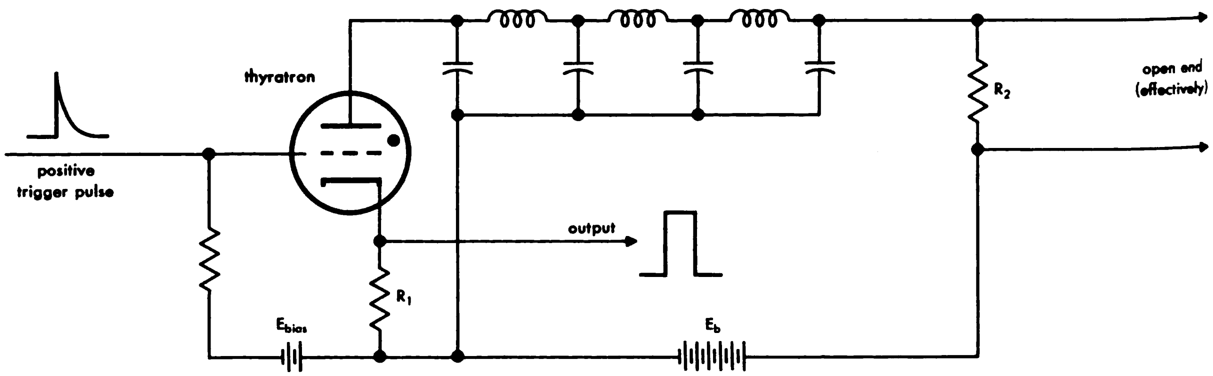
An artificial transmission line is used to create the modulating pulse either at a low-voltage power level (after which it is amplified by the modulator system) or directly at a high power level. The figure above illustrates a pulse-forming line with a mechanical switch.

The four condensers of the line are charged up to the potential of "E_b" through a high resistance (R_2). "R_2" has a resistance so much greater than the characteristic impedance (Zo) of the line that this end of the line may be considered effectively as an open circuit.

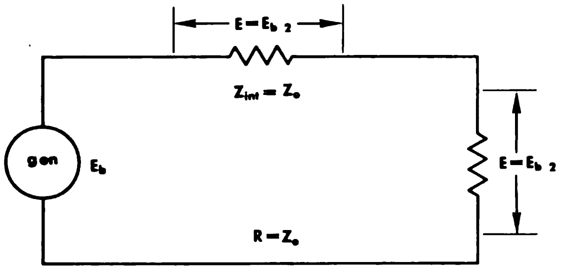
When switch "S" is closed, the line abruptly begins discharging at a constant rate through "R_1," which is equal to the characteristic impedance of the line. If the network consisted of the four condensers only, the discharge would follow an exponential curve, and the current flow through "R_1" would not be constant. However, by using a well-designed artificial line which contains both inductance and capacitance, the discharge rate can be held to a substantially constant value as long as the line is discharging.

At the instant the switch is closed, the discharge wave starts traveling down the line toward the open end, causing the wave to lose half of its charge. Upon arriving at the open end, the wave is reflected back toward "R_1." When it reaches "R_1," the line is completely discharged, and the current flow through "R_1" abruptly ceases, terminating the rectangular voltage pulse appearing across that resistance. When the switch is open, the line again charges to the potential of "E_b."

A more practical version of the foregoing circuit is shown below. Note that the me-



Pulse-forming line with a thyratron switch



Half of the line voltage is developed across the output in a pulse-forming circuit with a thyratron

chanical switch (S) of the previous circuit has been replaced by a thyratron tube, which permits the microsecond timing required in a radar system. The gas tube is normally nonconducting because of the negative bias on its grid. This allows the line to charge up to the value of "E_b" through "R_2." The line is switched to its terminating resistance "R_1" when a positive trigger pulse drives the gas tube into conduction. The current for this conduction is furnished by the line at a constant rate causing a rectangular wave of voltage to appear across "R_1" as long as the line discharges.

As the line becomes completely discharged, the plate voltage of the thyratron falls below the ionization point. As a result, the tube becomes nonconductive because, in a short period of time, "E_b" is unable to supply enough current through the high resistance of "R_2" to maintain the plate voltage. When the thyratron ceases to conduct, the line again charges up to the potential of "E_b." A relatively long charging time can be allowed since it normally takes place during the interval between transmitted pulses.

You can consider the charged line as a generator with an internal EMF of "E_b" and an internal impedance of "Zo." At the time the thyratron fires, the line (or generator) is effectively thrown across resistance "R_1," which is equal to "Zo" and in series with it. This equivalent circuit is shown above.

A voltage equal to one-half "E_b" appears across "R_1," and at the same instant a voltage wave "$-E_b/2$" starts down the line. This wave is reflected back from the open end in phase and arrives back at "R_1," having completely discharged the line. At this time, the thyratron ceases to conduct, and the voltage across "R_1" drops abruptly to zero. Thus, the potential has been maintained across resistor "R_1" during the time required for the voltage wave to travel down the line *and return*. You can see that the thyratron merely starts the action and that the time of its conduction is determined by the characteristics of the pulse-forming line.

It was stated earlier that for a delay line, the time delay is equal to $N\sqrt{LC}$. However, in the case of a pulse-forming line charged to a given potential, the voltage wave must move down the line and back, thus producing a time delay twice as great. Thus, the width of the voltage pulse developed across the terminating resistance is determined by:

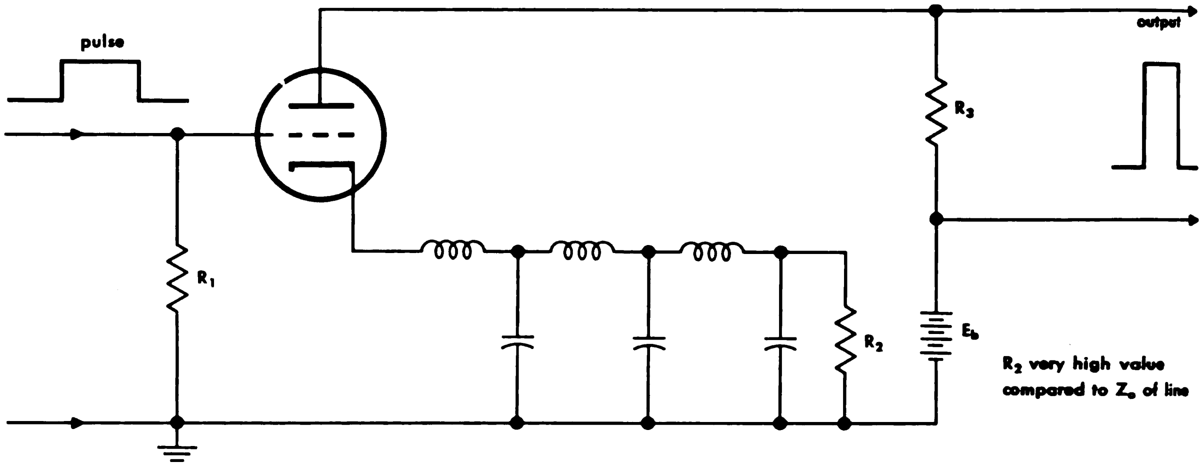
$$T=2\ N\sqrt{LC},$$

while the pulse repetition rate is dependent on the frequency of the timing oscillator which triggers the thyratron switch tube.

A variation of the pulse-forming circuit is shown on the next page. In the circuit, the pulse-forming line is in the cathode circuit of a triode amplifier tube.

Let's assume that the amplifier tube is normally nonconducting and that it is desired that the tube produce a rectangular pulse of short duration at the output. The cathode resistor of the amplifier is replaced by an artificial line, and the tube is brought into a conducting state by the sharp rise in voltage as the leading edge of a rectangular wave is applied to its grid. At this instant, plate current increases sharply and establishes a voltage wave between cathode and ground. This voltage starts down the line, and upon reaching the *effectively open end* (R_2), it is reflected back without change in polarity, arriving between cathode and ground after a time "$2N\sqrt{LC}$." This sudden increase in voltage on the cathode is sufficient to cut off the tube. Thus, the *leading* edge of the output pulse (appearing across R_3) is formed as the tube is placed in conduction and coincides with the leading edge of the pulse on the grid.

The constant amplitude of the output pulse is maintained by the characteristics of the line. The *trailing* edge of the output pulse is formed as the line cuts off the tube at the

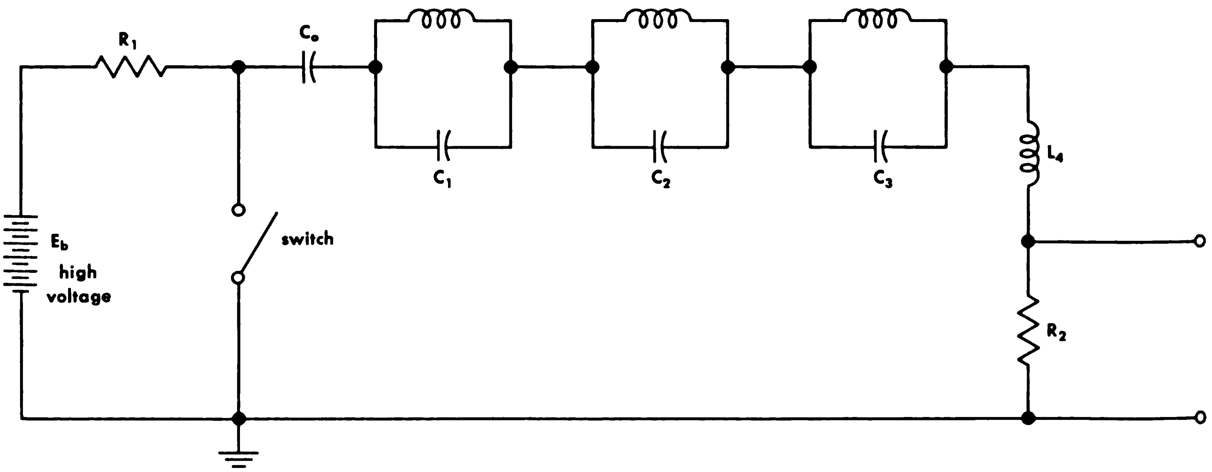


Pulse-forming line with a thyratron switch

time "3 N$\sqrt{LC}$" and bears no relationship to the grid voltage. After the tube is cut off, the line discharges slowly through "R_2" until the cathode voltage becomes low enough to allow the tube to conduct, or until a pulse again drives the tube into conduction, initiating a new cycle of operation.

The foregoing circuit does not necessarily have to be an amplifier stage; it could be an RF oscillator, or transmitting tube, with an artificial line in its cathode circuit. Such an application would determine the width of the RF pulse transmitted whenever the transmitter circuit is triggered by the timing oscillator.

While the standard artificial line is quite satisfactory to produce a time delay, it has some limitations as a pulse-forming device due to the fact that each section of the line tends to produce a hump, or ripple, in the flat top of the rectangular-wave ouput. This defect can be minimized by using a large number of small units to obtain the required amount of inductance and capacitance. The resulting humps, or ripples, while more numerous, would be of much smaller amplitude. This would result in an artificial line more closely approximating a real transmission line. However, the artificial line composed of a large number of small LC units is less compact and is more complex in design.



Guillemin line with mechanical switch

Actually, when pulses are formed by artificial lines at low power levels and then amplified, some deviation in the pulse shape can be tolerated. Limited deviation is permissible because amplifiers can be operated so as to effect a small mount of plate limiting and thus clip the uneven flat top before it is applied to the transmitter tube.

When pulses are to be developed initially at high power levels for direct application to the transmitter tube, little can be done to improve the pulse shape when using a standard artificial line as the pulse-forming medium. To meet the rigid pulse-forming requirements of this latter application, a special type of line, known as the *Guillemin line*, has been designed. This line is shown in the diagram below left as employed with a mechanical switch. The line is charged through "R_1," which could be replaced by a choke. "C_o," the main storage capacitor of the line, must withstand the full potential (E_b). Closing the switch causes the line to discharge through "R_2." The line is so designed that on discharge a more nearly constant current flows through the terminating impedance than can be obtained with the standard type of line. A rectangular wave with a more perfect flat top is obtained from the Guillemin line, and it may be used to create pulses of high amplitude for direct application to the transmitting tube.

In practical application, the switch is replaced by one or more thyratrons with their cathode-plate paths in series so as to divide the high voltage of "E_b." The discharge of the line is effected by triggering the grids of the thyratrons. "R_2," in this case, would be replaced by the transmitting tube. And the impedance of the line, as well as the impedance of the transmitting tube, would be matched to prevent reflections.

There exist many modifications of the foregoing pulse-forming circuits, as well as other devices which are employed to produce pulses. However, the circuits presented here are basic circuits and are found in general use in microwave radar as used in conjunction with missiles.

Programmers

A flight programmer is a unit into which precalculated flight data is inserted. It supplies this data to the various control circuits of an aircraft in the sequence necessary to cause the aircraft to conform to the predetermined pattern of flight. The programmer may include such units as the air log, Veeder counter, map-matching devices, magnetic-tape pickoffs, and stabilized platforms. These units may be found singly or in various combinations in the programmer, depending on the type of control required or the complexity of the flight pattern.

AIR LOG. An air log is used to determine the range of flight. It operates on the principle of an air screw which makes a specific number of revolutions while moving through air for a given distance and at a given velocity. The number of revolutions per unit of distance depends on the pitch of the screw and the density of the air.

Generally, an air log is attached to the outer surface of the nose of the aircraft and consists of a small four-bladed propeller, or screw, mounted on a shaft which drives a reduction worm gear with a ratio of 30 to 1; that is, for every thirty revolutions of the air screw, the driven gear makes 1 revolution.

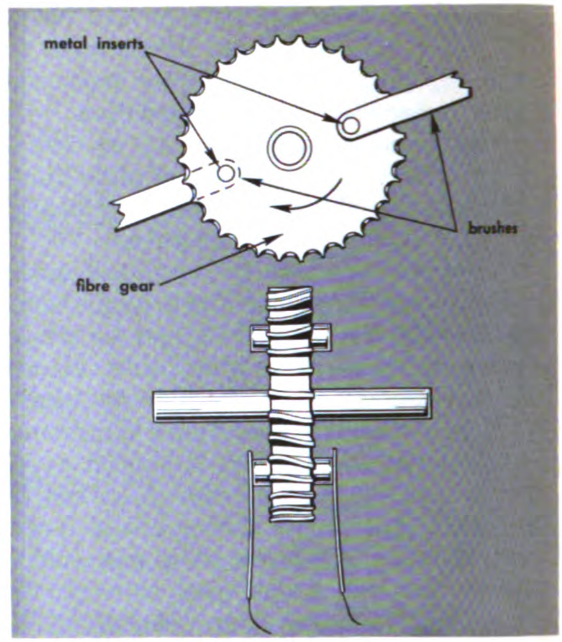
The driven gear is machined from hard fibre, or some suitable dielectric material, into which is inserted a pair of metallic contact points which make electrical contact with brushes. Two impulses are transmitted for each revolution of the gear. Thus, one impulse is transmitted for every fifteen revolutions of the air screw.

The contact points and brushes serve to close a magnetic relay circuit in a Veeder counter, once for each 15 revolutions of the air log.

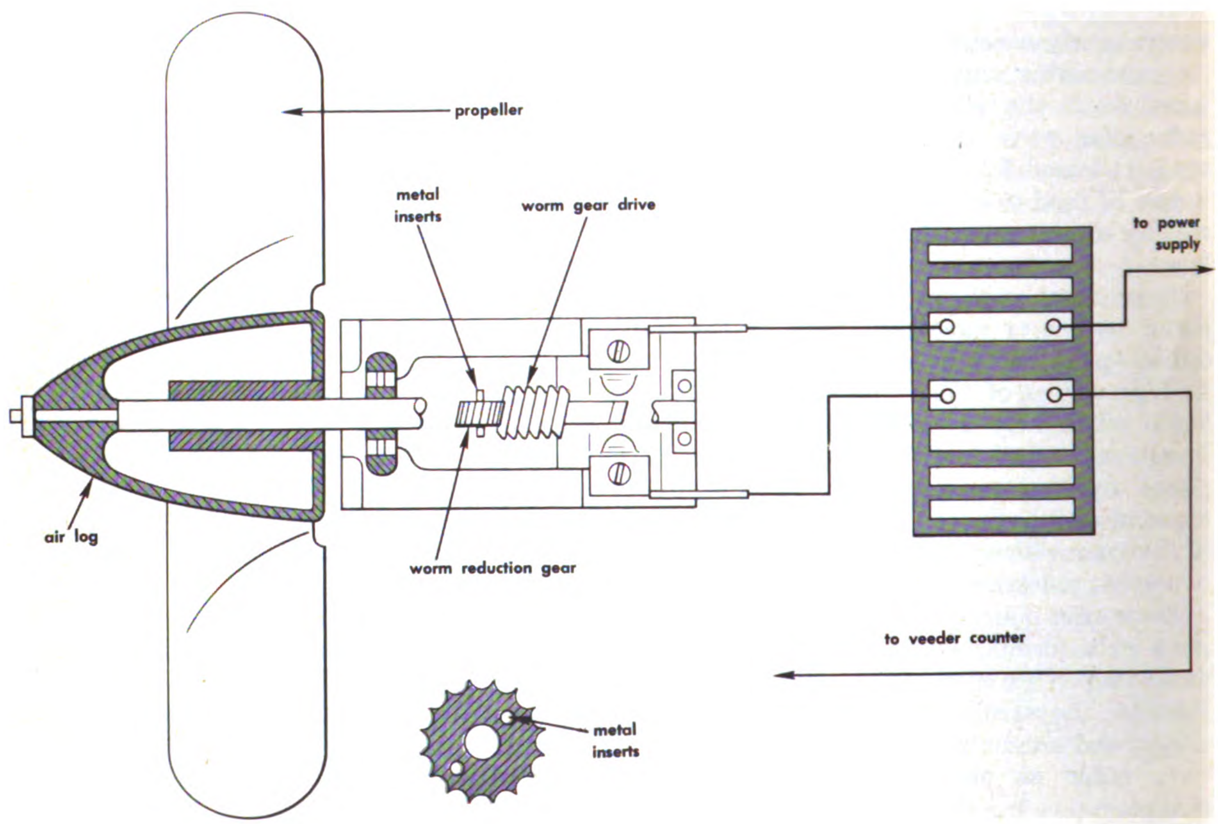
Basic structural details of an air log are illustrated in the following sketches.

VEEDER COUNTER. A Veeder counter is a device consisting of several calibrated drums. The drums are driven by a train of gears designed to provide the desired ratio of turns between the drums. The mechanism is similar to that found in the odometer unit of an automobile speedometer. The calibrated drums are rotated and set to the desired length of travel for the missile. Each impulse from the air-log-operated relay represents a specific distance traveled (a specific number of revolutions of

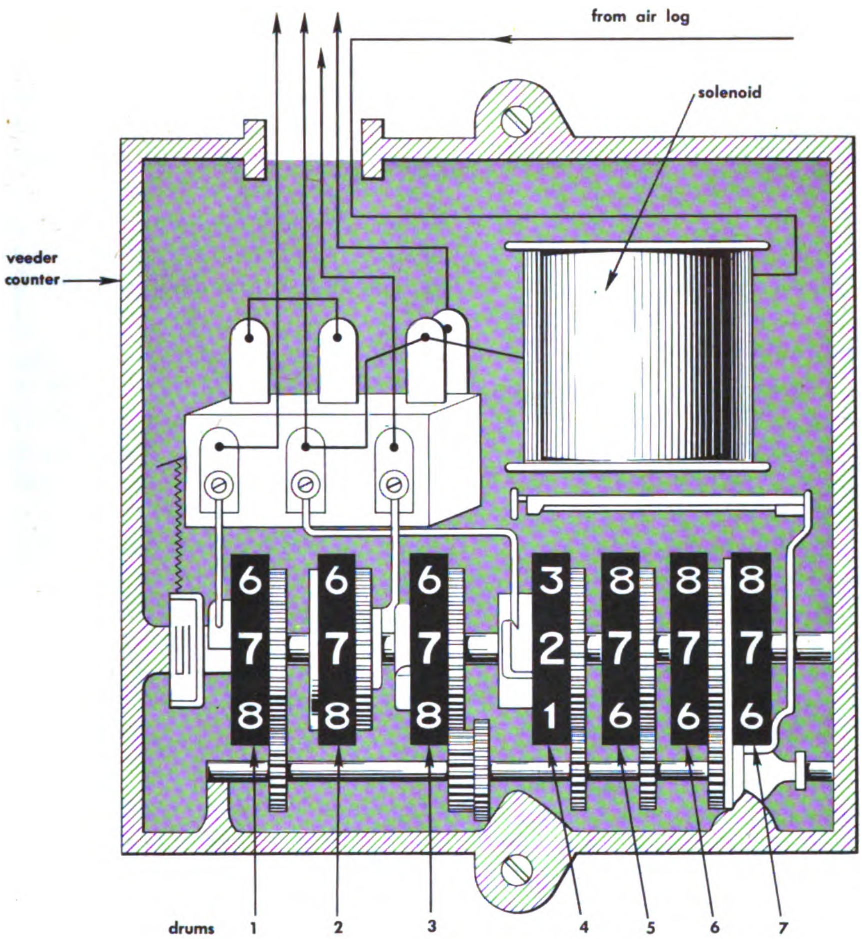
the air screw). Each one actuates the Veeder-counter mechanism so that the calibrated drums rotate from their preset position back toward zero. When the counter reads zero, it indicates that the predetermined dump point and/or arming point has been reached. At zero reading, the proper circuits are energized to position the controls for the terminal dive and to arm the warhead.



MAGNETIC-TAPE PICKOFF. The need for a flight programmer capable of storing sufficient data for long-range flights, and having a degree of accuracy sufficient to render it suitable for use in high-velocity missile, has led to much experimentation with magnetic tape and wire data recording and playback units. These devices make it possible for predetermined flight data to be magnetically recorded on tape or wire and used as a reference in guidance of missiles.



Structural details of an air log



Cross-section of veeder counter

In celestial-navigation systems, predetermined star-position data, such as desired azimuth and elevation angles with respect to the position of the missile at all points or at any specific points along the flight path, are compared with the indicated position data from the star tracker. Any error between the tape data and tracker data is utilized to reposition the missile back onto the precalculated course. This necessitates exact recording playback of data, which in turn requires a timing and synchronizing system with a high degree of accuracy and stability.

When employed in a celestial-navigation system, the tape data specifies that at every point along the flight path the angles between the missile and the reference stars should be of some exact value. If the star tracker indicates a discrepancy between the precalculated angle data and the actual position of the missile, the gyros, torquers, and accelerometers on the stabilized platform feed information into the tracking control circuits to bring the missile into the correct angular relationship with the reference stars.

In the trajectory playback system, a time standard and a magnetic-tape playback reproduces in the missile information which has been calculated and recorded in advance. This information controls the various circuits of the guidance system according to a predetermined schedule.

Fine steel wire affords a medium for storing a great amount of data on a spool of small size and is relatively indestructible. However, certain disadvantages make wire recording and playback unsuitable for flight programming. These disadvantages include inaccuracy in sequencing due to slippage of the wire on the spool or reel and nonuniform wire speed re-

sulting from variations in the rate of unwinding. As the wire unwinds, the diameter of the storage spool decreases while that of the pulling (capstan) spool increases. This operation gradually increases the rate at which the wire unwinds. No suitable device has been developed for pulling the wire at a constant speed independently of the reels. This disadvantage of variation in wire speed prohibits the use of wire in large spools as would be required for programming long flights.

Positive, uniform recording and playback speeds can be more readily obtained by the use of punched tape of film. Tape and film strips can be designed to provide several channels of data simultaneously. When steel tape or non-shrinkable film is used, and more than one set of data is impressed upon the tape, it is possible to store much more data with less bulk on tape than on wire. However, multichannel tape requires the use of a multiple magnetic pickoff head with separate amplifiers and filters for each data channel.

The basic principles involved in magnetic tape and wire recording are the same; namely, the data impressed by electrical impulses is recorded on the tape or wire in the form of magnetized areas. Information is represented by the degree of magnetization in each unit section of tape or wire, or the magnetized areas are coded by spacing or by number with respect to reference points on the tape.

With any flight programmer, it is essential that a highly accurate timing system be employed so that the sequencing of the flight-control data shall coincide exactly with the speed and flight position of the aircraft. In some magnetic-tape programmers, a crystal-controlled oscillator is used as a master time-base generator, and its frequency is stepped down through a series of Eccles-Jordan multivibrators (note the figure on the right) to produce the desired time-base frequencies. This method requires a large number of stages, which increases the likelihood of error due to tube or component failure or to fluctuations in power-supply voltage, temperature, etc.

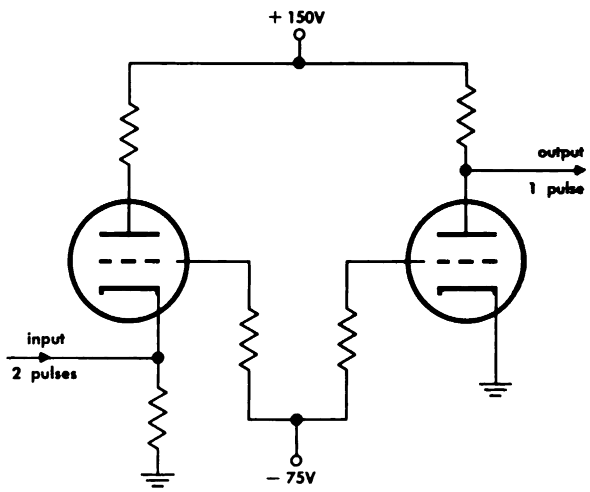
Time-base generators or oscillators employing tuning-fork control at low frequencies are more accurate and stable than those of other types. They are becoming more widely used in flight programmers. The accompanying figure is a functional block diagram of a magnetic-tape programmer which employs a multichannel tape with each channel capable of carrying two sets of data simultaneously. One set of data is recorded by magnetizing pulses of one frequency, such as 200 cps, and the other set is recorded by magnetizing pulses of a different frequency, such as 400 cps.

The tape is passed through a multiple pickoff playback head of the magnetic or variable-reluctance type with one pickoff for each data channel. Dual frequency outputs from each pickoff are amplified and passed through two filter circuits, one of which rejects the 200-cps data and passes the 400-cps data, while the other rejects the 400-cps data and passes the 200-cps data. The data pulses from each channel filter are amplified further and then applied to relay circuits for actuating the controls.

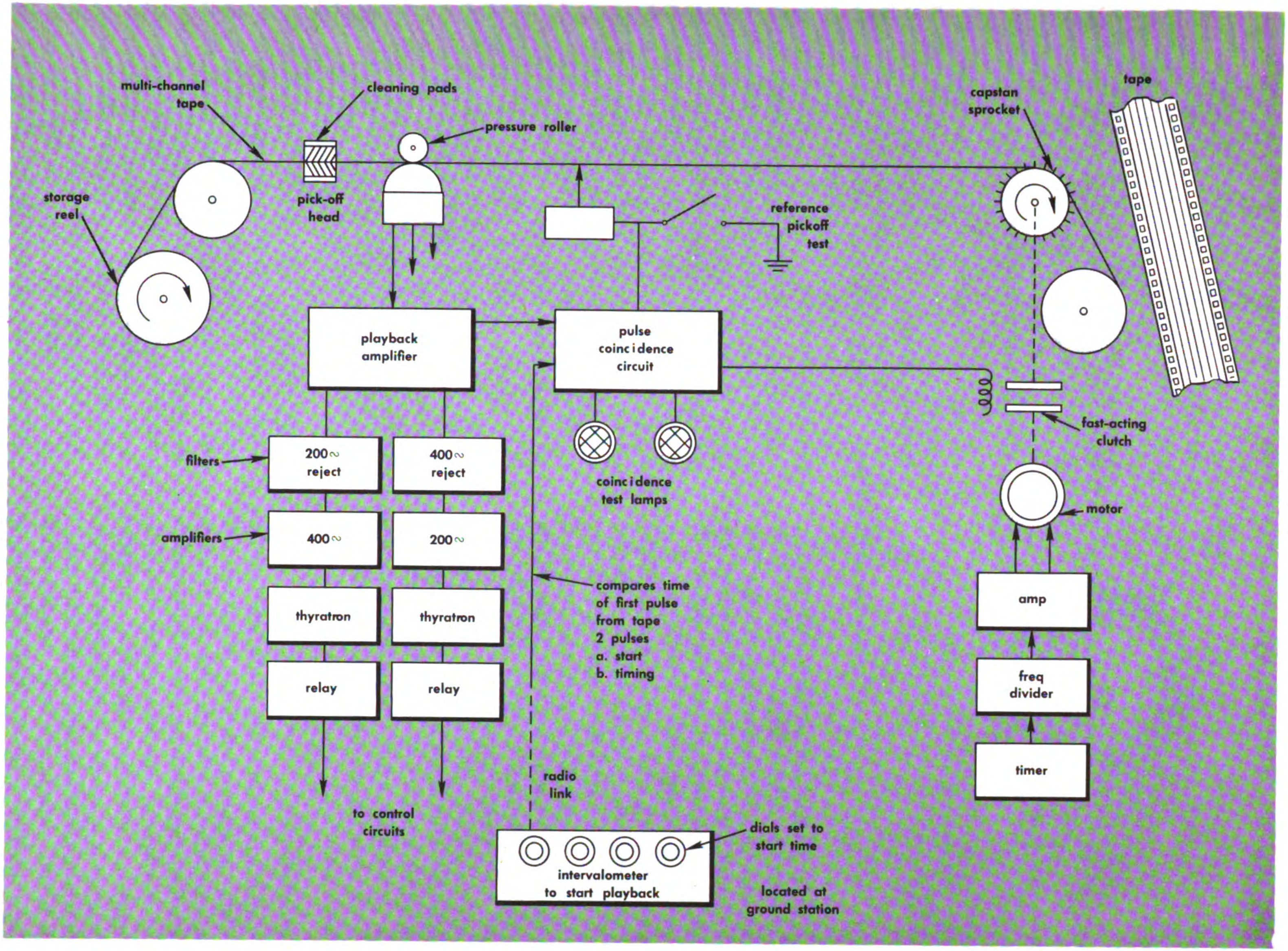
In the block diagram, a thyratron stage is shown in each channel. The purpose of this stage is to conserve power by applying current to the relays only when a data signal is present.

The functions of the other blocks are shown on the diagram. The individual circuitry is treated in detail throughout this Chapter, beginning with Section I.

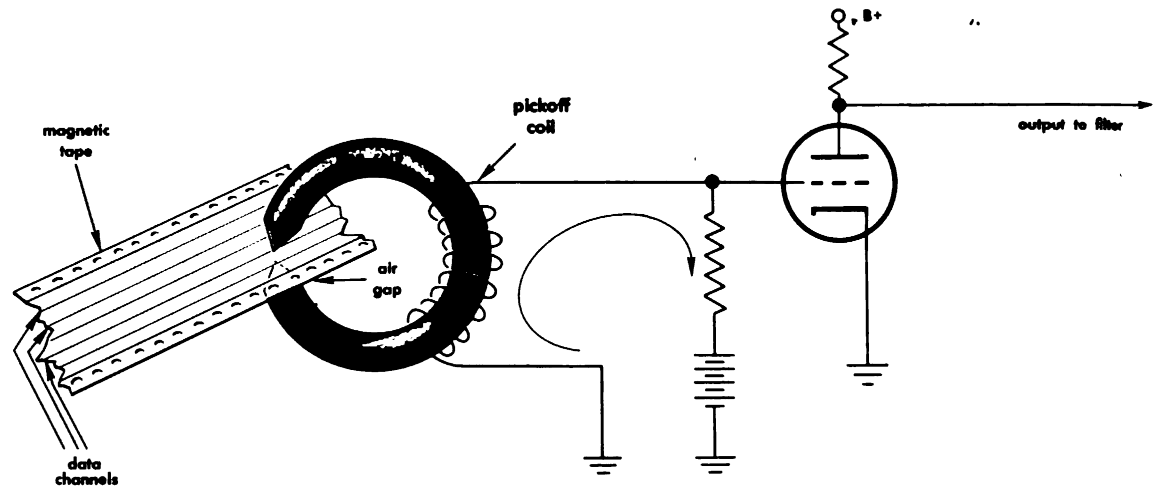
The magnetic pickoffs operate on the variable-reluctance principle, in which the iron



Eccles-Jordan multivibrator used as a step-down counter



Magnetic tape programmer



Variable-reluctance magnetic pickoff

core of the pickoff coil is designed with a small air gap to offer a high-reluctance path for magnetic flux. The tape passes through this gap and varies the reluctance of the core by an amount proportional to the degree of magnetization of the data channel, or at a rate proportional to the number of magnetized sections passing through the gap in a specific unit of time. When the reluctance of the core is varied, the same effect is produced as would be obtained by varying the inductance of the coil, and the current flowing through the coil will vary accordingly. These variations in current constitute the data or intelligence conveyed from the tape to the controls through the playback circuitry of the programmer.

The preceding figure represents the design principles of a magnetic (variable-reluctance) pickoff.

Map Matching

Radar map matching is another method for providing guidance reference in preset or automatic navigation of missiles. In this system the terrain which lies beneath the craft's course is scanned by radar, and the PPI-scope presentation thus obtained is compared with a map or photograph of the same terrain. Any variations between the map and the PPI presentation, after conversion to the same scale, are resolved into error voltages that are used to reposition the aircraft in the manner required to eliminate the error.

Stabilized Platforms

The accuracy of a missile guidance system is dependent on the accuracy and sensitivity of its individual components and its ability to compensate for the effects of external forces acting on the aircraft. To maintain a predetermined relationship between the missile and a fixed reference at all points along the flight course requires the resolution of all forces tending to produce any deviation from the desired pattern of flight.

In inertia guidance systems and in automatic celestial navigation systems, an artificial horizon plane always maintains a position perpendicular to the normal flight path. This plane provides an accurate reference for measuring star angles and determining instantaneous position of the aircraft. This artificial horizon plane incorporated into the aircraft is referred to as a *stabilized platform.*

A stabilized platform is necessary in all long-range guidance systems. It provides the reference plane for a predetermined path where the path of the missile is adjusted by devices *wholly within* itself. It makes use of Newton's laws of motion. The platform is independent of outside information except in the celestial type systems where the telescopes are used to help keep the platform stabilized and oriented. In order to resolve all of the external forces acting on the missile into either roll, pitch, or yaw errors, the stabilized platform

must be kept perpendicular to the direction of the local gravity vertical at all times. The reason for this is that unless the platform is maintained perpendicular to gravity, a horizontal component of gravity will act on the sensing instruments on the platform indicating false accelerations.

This horizontal component is shown as g sin a in the diagram on the right. Alpha (a) is the angle between the normal to the platform and the local gravity vertical or the direction of gravity at that geographical position.

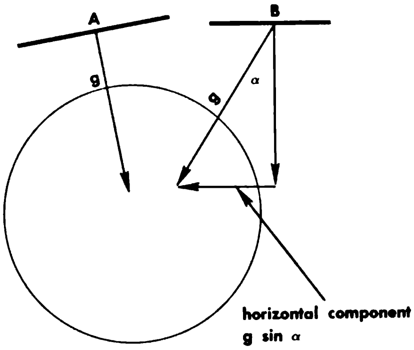
In space, a body in motion continues in motion in a straight line with a constant velocity, unless acted upon by some external force. Accurate measurement of the degree of such external forces, with correction factors, would enable a missile to hit any target within its fuel range.

For detection of any deviation of a missile from a predetermined path, the objective or target must be stationary. The target must retain fixed position with respect to the guidance reference at all times or reference is of no value. The path is initially defined by a platform stabilized by gyroscopes around roll, pitch, and yaw axes. These gyros are given a space reference and would take the missile out into space unless corrected. To make the missile follow a path around the earth, it is necessary to torque the gyros so that they precess and orient the platform in its proper position as the missile follows the curvature of the earth.

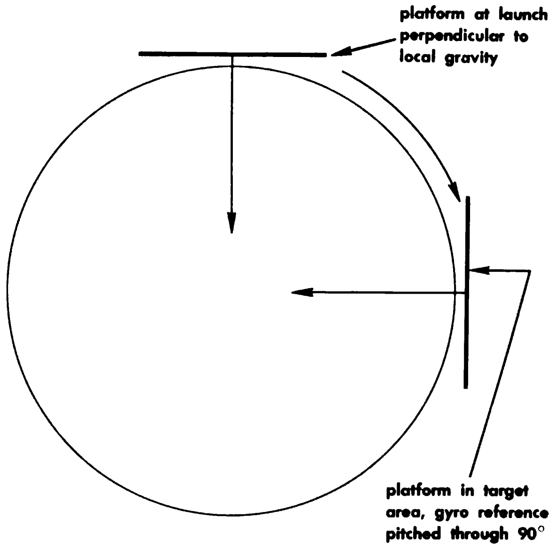
If the missile were to fly to a target one-half way around the earth, the platform reference would be pitched through 80° from its original position!

To determine the amount the platform is pitched or precessed, position of the missile must be known. Position can be determined by knowing the velocity and time of flight. To detect any change in velocity, an acceleration detecting device is added to the platform. The velocity change could be noted by a device whose action is similar to that of a plumb bob. A displacement of the plumb bob would indicate a change in velocity, but in any constant condition it would hang vertically. A change in velocity *is* acceleration. If this acceleration is integrated and the velocity and distance outputs compared with the desired distance and velocity, the position of the missile is easily determined. The functions of distance, velocity, and direction are then set into a program device. This program device specifies a velocity, and any external force acting on the missile produces a change in velocity or an acceleration.

This concludes our discussion of reference units. We take up amplifier units of guidance systems next.



Platform improperly precessed



Platform correctly precessed

Amplifier, Controller, Actuator, and Feedback Units of Guidance Systems

This section covers amplifiers, controller, actuator, and feedback units of missile guidance systems. These units are grouped here in one section not because they are less important than other guidance units but because of the brief treatment given them in this text. You must keep in mind that these four units are as vital to effective missile guidance as the units already discussed at greater length.

Let's first consider amplifier units.

AMPLIFIER UNITS

Certain operations of amplification may appear in any of the units of the block diagram shown below. The discussion here, however, covers various types of amplifiers not previously discussed.

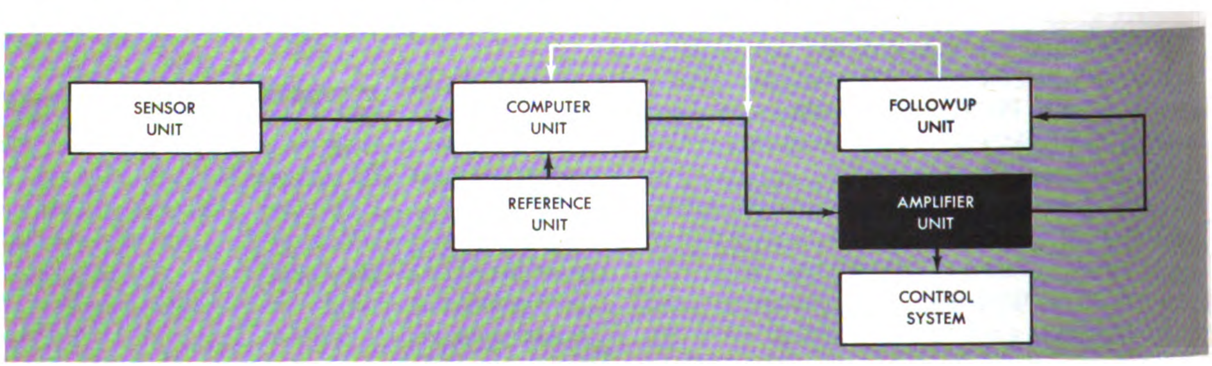
Frequency Selective Amplifiers

Frequency selective amplifiers are employed in FM receivers which are used in guided missiles to receive control signals from control transmitters. The control transmitters are located on the ground or in a mother aircraft.

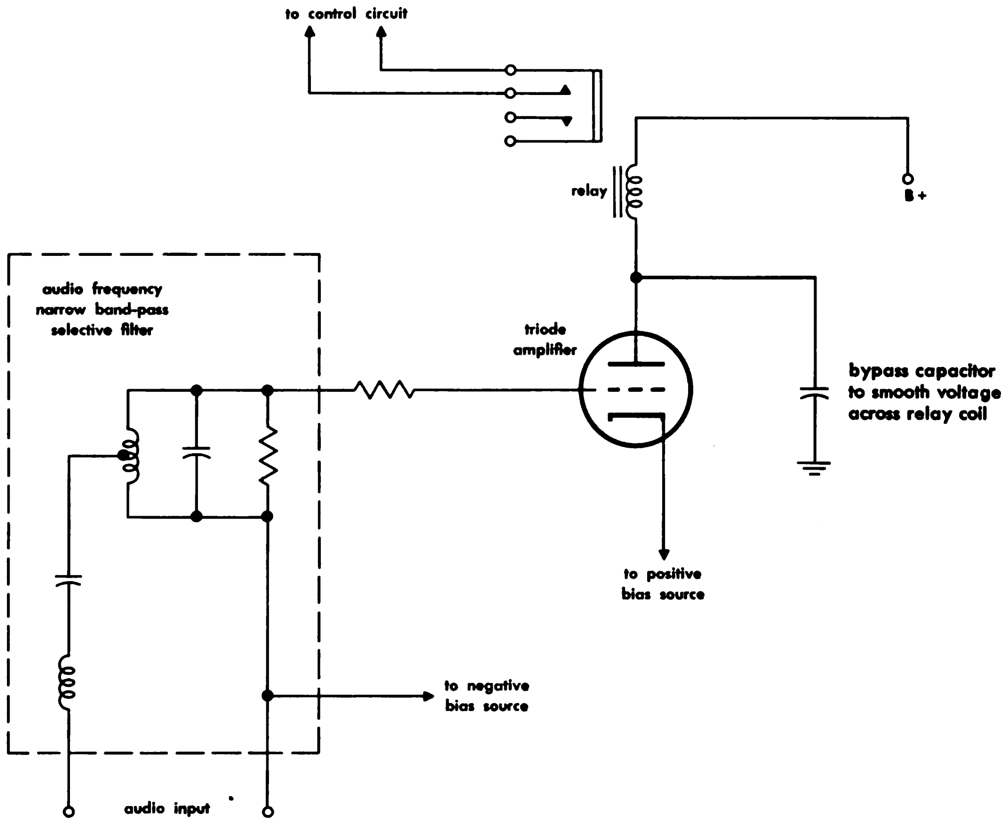
The control signals are frequency modulated by one or more of a number of audio channels; the number of channels varies with the requirements of the system. These signals are utilized to control servomechanisms, auxiliary functions, and safety equipment in missiles.

The audio control frequencies, after being detected, are routed by selective audio filters to relay energizing circuits which cause relays to close or open, energizing or deenergizing the synchro units associated with the control mechanisms. Each audio channel, comprising a relatively narrow band of frequencies, is used to actuate a specific control in the missile.

The diagram at the right illustrates a typical frequency-selective amplifier stage with audio filter and relay control. This is one form of a tuned-grid amplifier.



Basic missile guidance diagram



Frequency-selective amplifier

The filter section is a combination of a series resonant circuit and a parallel resonant circuit with the constants so chosen that a relatively flat and narrow audio bandpass is obtained. A tapped input in the coil of the parallel resonant circuit gives the filter an overall voltage gain at resonance.

Under *no signal* conditions, the positive bias on the cathode and the negative bias on the grid keep the tube cut off. The relay in the plate circuit is not energized. When an audio signal is received, it is increased in voltage by the action of the parallel resonant section of the filter. It then is applied to the grid of the triode tube. The high signal voltage on the grid of the triode overcomes the bias and causes the tube to conduct heavily. The plate current energizes the solenoid of the relay, causing the relay contacts to close.

A pulsating voltage tends to exist across the relay coil because of tube cutoff during the negative swing of the audio signal cycle and because of the high plate current during the positive swing. This pulsating voltage is smoothed out by the filtering action of the plate bypass condenser. The filtering action makes the current through the relay solenoid relatively steady while the audio signal is being received. Several of these frequency selective amplifiers may operate simultaneously in a well-designed receiver.

Other amplifier circuits employed for frequency selection include:

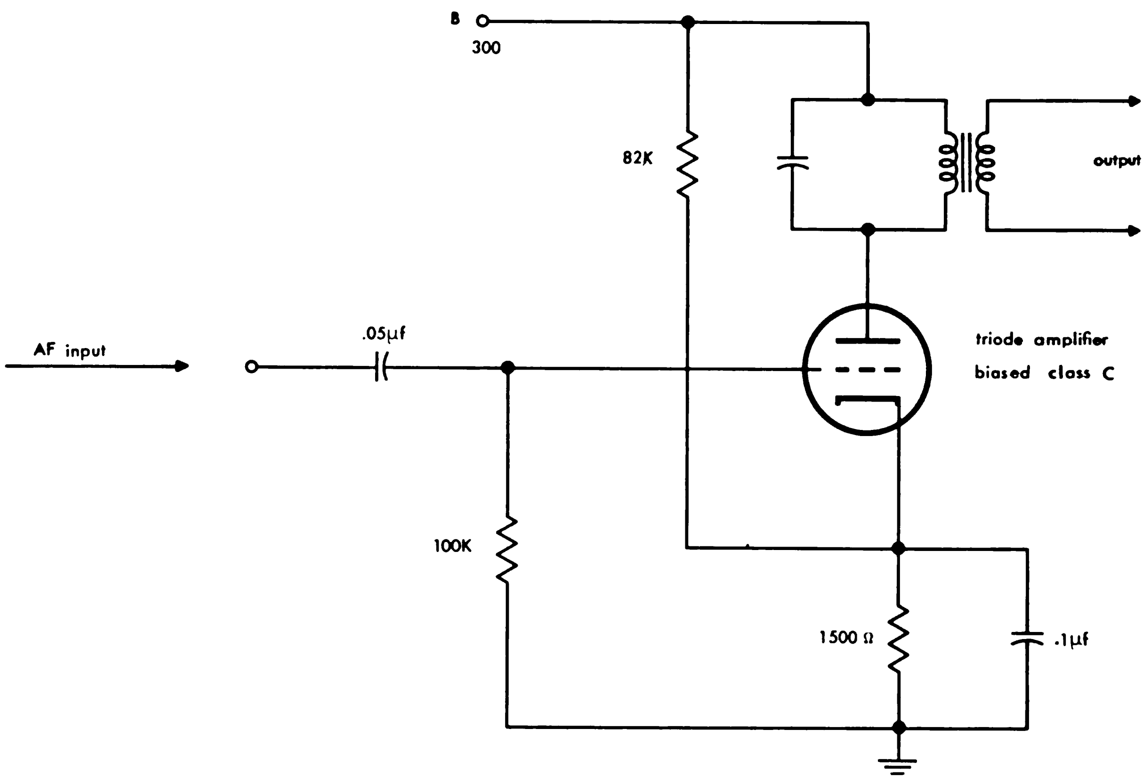
a. Tuned plate-frequency selective.

b. Tuned plate-tuned grid.

c. Tuned cathode.

Each of these circuits is especially suitable for some specific application, but the primary function is the same; that is, their function is the selection and amplification of some particular band of audio frequencies.

Tuned plate-frequency selective amplifier. A tuned plate-frequency selective



Tuned-plate frequency-selective amplifier

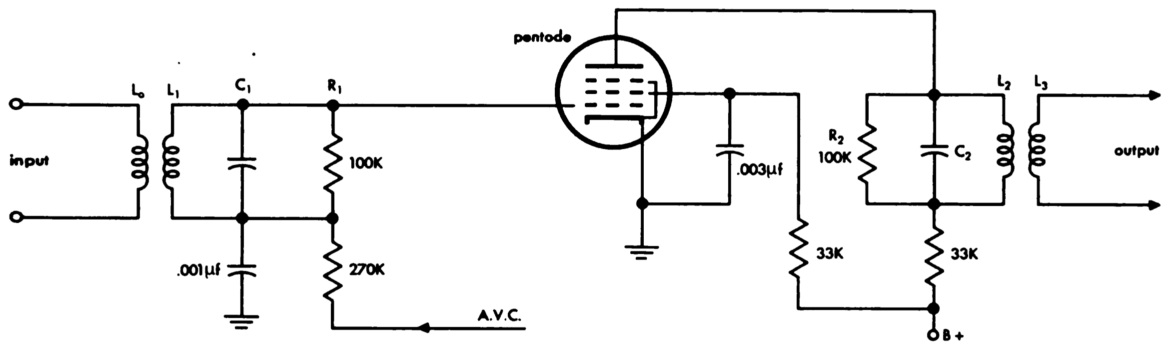
amplifier is illustrated in the accompanying diagram. This circuit represents a class "C" amplifier with its plate tank circuit tuned to resonance at the frequency desired for amplification. Typical values for components and plate supply voltages are shown in the diagram

A positive voltage is applied to the cathode, for class "C" bias, through the voltage divider consisting of the 82-K and 1500-ohm resistors. The audio signal is applied to the grid through the .05-mf coupling capacitor and developed across the 100-K grid-leak resistor.

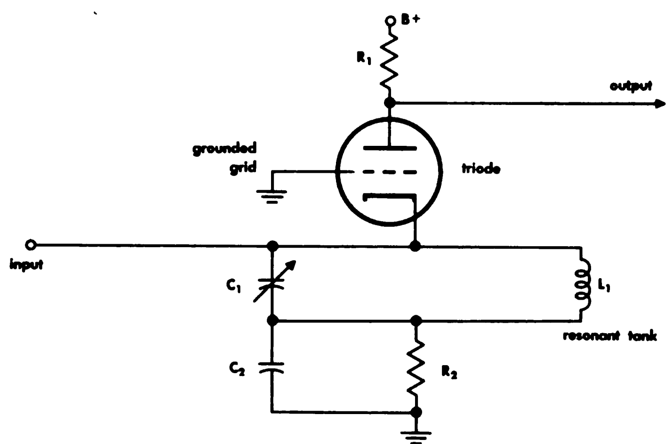
The output signal is inductively coupled to the next stage. The plate tank circuit, being tuned to resonance at the desired audio frequency, selects the frequency to be coupled to the succeeding stage. Amplification results from the action of the tuned circuit, the "Mu" of the tube, and the ratio of the transformer coupling.

TUNED PLATE-TUNED GRID. The diagram at top right illustrates a frequency selective amplifier employing a tuned plate-tuned grid circuit. The tanks use variable capacitors for tuning. The 100-K resistors across the tank circuits lower their "Q" and thus widen their bandpass. The .001-mf capacitor in the grid tank circuit keeps the grid below ground potential for automatic volume control (AVC). This circuit is most widely used in IF stages using IF transformers to couple the signal from one stage to the next. RF energy coupled from the input coil is built up in voltage in the tuned circuit, $L_1C_1R_1$. The voltage that appears across "L_1C_1" is applied to the grid of the pentode which further amplifies it. A second resonant circuit, $L_2C_2R_2$, is the load for the plate of the tube and offers high impedance because it is a parallel resonant circuit. The output is inductively coupled to "L_3." Most pentode tubes are designed to operate with the screen grid at a lower potential than the plate. A bypass capacitor is used to keep the screen grid at a steady potential by passing to ground any signal that appears on the screen grid.

TUNED CATHODE. Another widely used frequency selective amplifier is the tuned cathode amplifier, shown on the right. This



Tuned grid-tuned-plate frequency-selective amplifier



Tuned cathode-grounded grid amplifier

frequency selective amplifier, which has become widely used in radar and television RF and input IF stages, has the tuned circuit in the cathode and uses a resistive plate load. The signal is applied to the cathode.

In this circuit, the gain of the stage is less than when the signal is applied to the control grid, but the loss of gain is compensated for by a reduction in noise level, so that a less distorted signal is transferred to the following stage.

"C_1" and "L_1" comprise the tuned circuit which amplifies only the frequency band to which it is tuned. "R_2" and "C_2" develop self-bias for the amplifier.

The above frequency selective amplifiers are found in the amplifying units of many missile guidance systems. It is essential that an amplifier having high·selectivity be used in the missile in order that precision control can be attained.

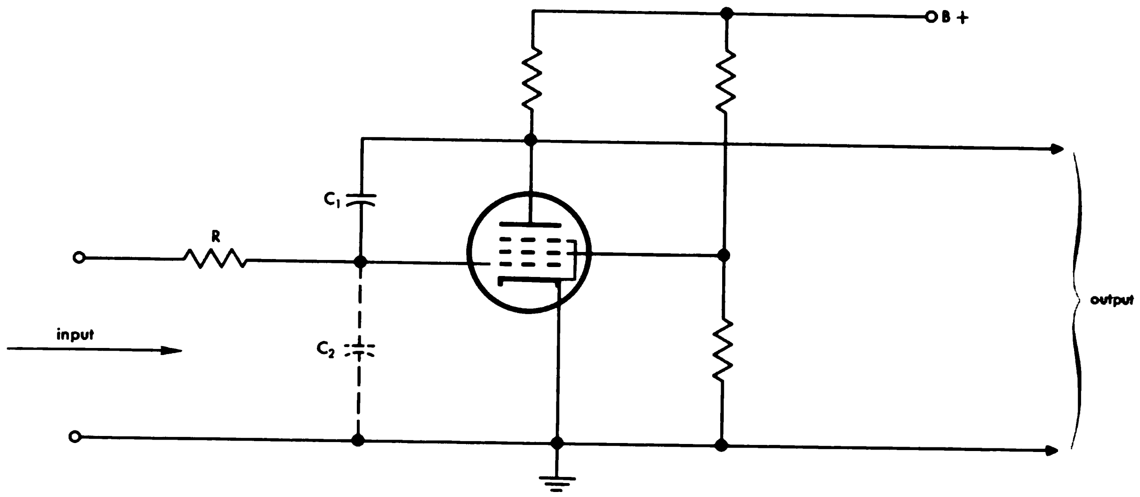
Miller Integrators

Miller integrators are found in many radar pulse shaping circuits and in high-fidelity amplifiers in which a large value of negative feedback is used to obtain linearity.

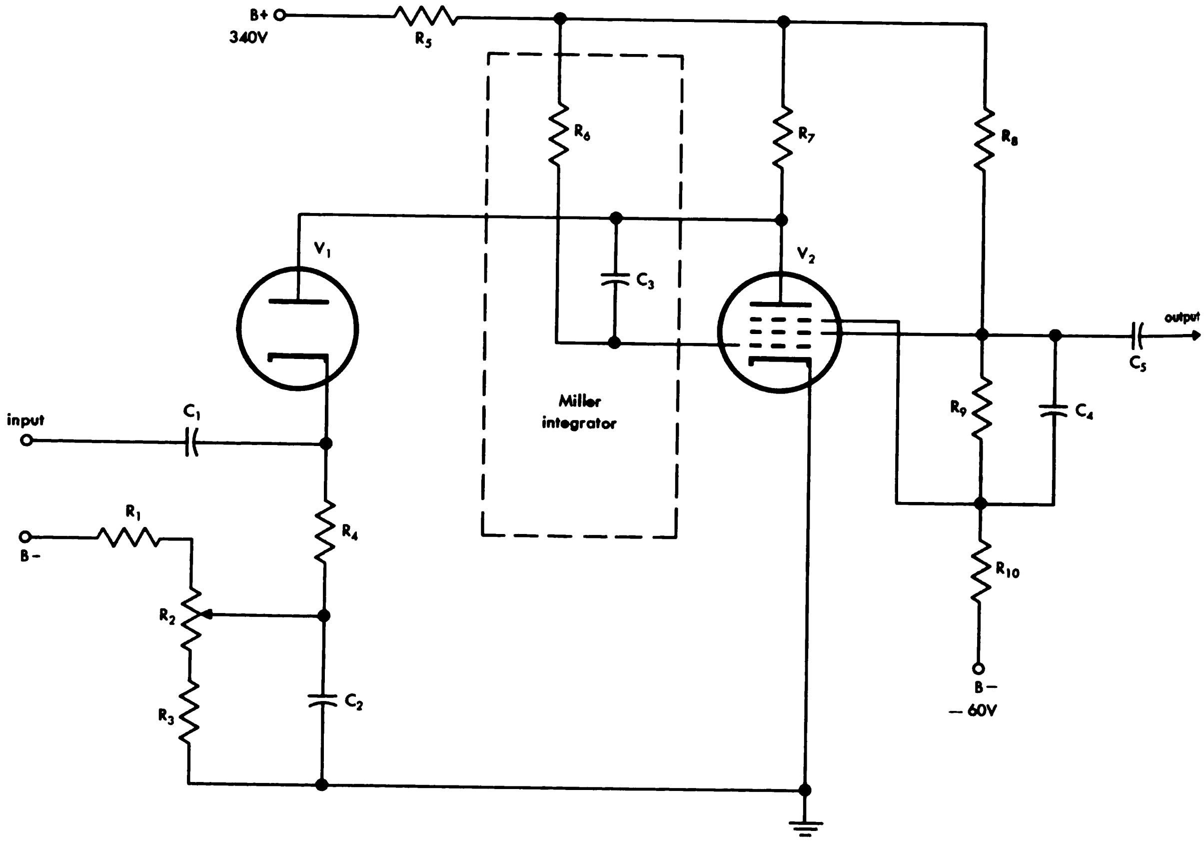
The Miller integrator shown on the next page makes use of the amplified value of grid to plate capacitance (C_1) to obtain a large effective value of input capacitance (C_2). This capacitance with the input resistance (R) comprises the integrating circuit.

A pulse-shaping phantastron circuit, as shown in the bottom figure on the next page, utilizes the Miller effect integrator circuitry, thus producing a positive pulse of variable width to code pulses of RF energy in pulse-width coded systems.

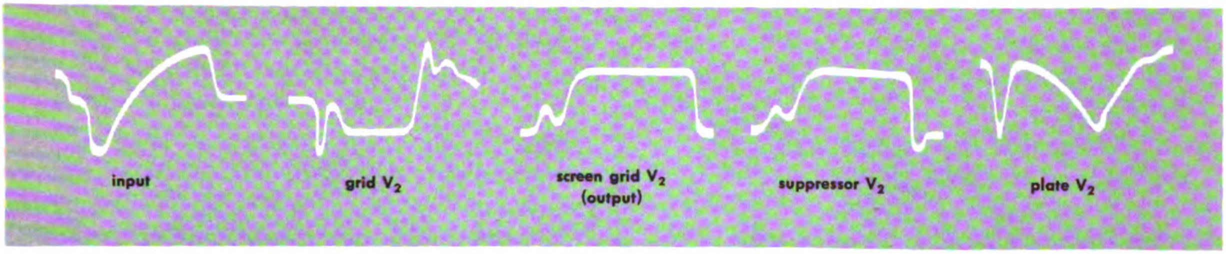
The width of the output pulse is controlled by the setting of potentiometer "R_2." Output is taken from the screen grid of "V_2" which



Simple Miller-effect integrator



Pulse-shaping circuit with Miller-effect integrator



Wave forms of pulse-shaping circuit with Miller-effect integrator

is a pentode connected in a *phantastron* circuit with its plate and grid connected by an RC network to function as a Miller integrator.

Input tube "V_1" receives the negative input pulse at its cathode through coupling capacitor "C_1." The signal leaves "V_1" as a negative trigger pulse and is applied to the grid of "V_2" through capacitor "C_3." This trigger pulse initiates one cycle of the phantastron action.

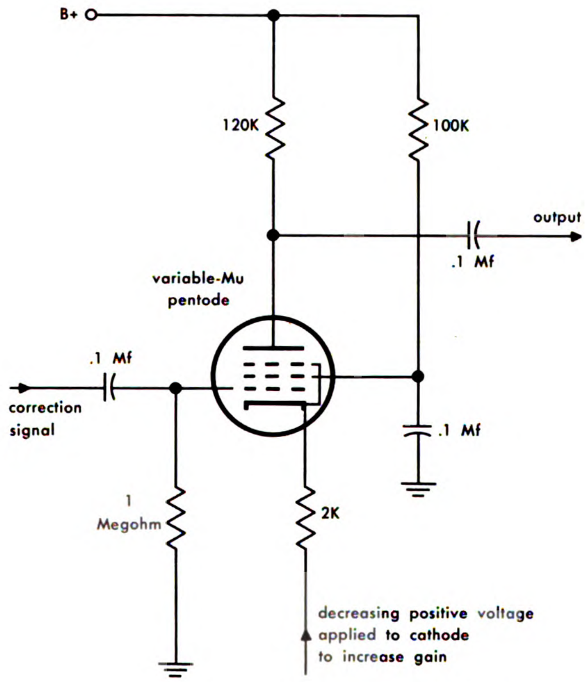
The wave forms shown in conjunction with the circuit diagram of a pulse shaping circuit illustrate how the circuit is capable of converting a narrow negative input pulse into a positive output pulse of symmetrical form and variable width.

Miller integrating networks are also found in many applications, such as DC amplifiers, in which it is desired to produce a change in the output only after a certain period of time.

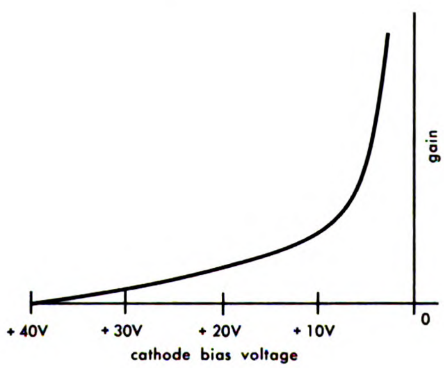
Mixing Amplifiers

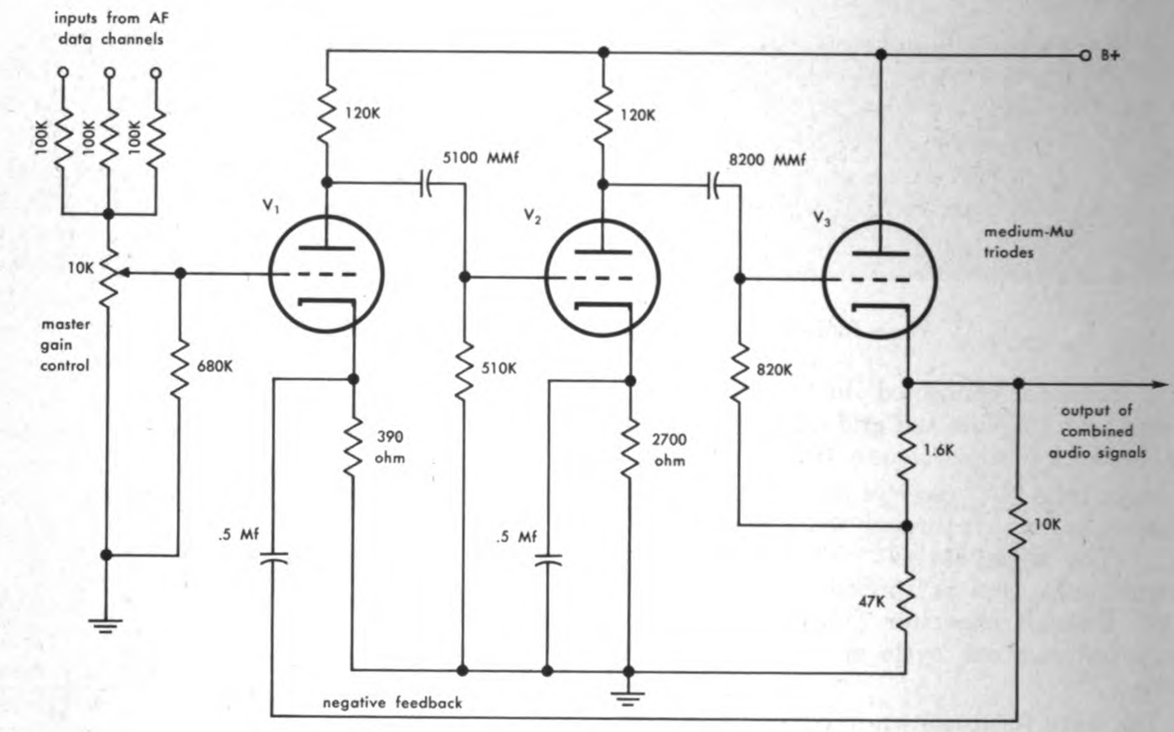
Mixing amplifiers are used whenever it is desired to impress AC or RF pulses upon a varying DC signal, as in telemetering systems. They are also used to control amplification of correction voltages transmitted to the guidance and/or control system of guided missiles when it is desired to increase sensitivity of the guidance system as the missile approaches the target.

The figure on the right illustrates a mixing amplifier circuit in which a decreasing positive voltage is applied to the cathode of a variable-Mu pentode. This produces the same effect as applying an increasing positive voltage on the grid. In either case the bias is decreased. Voltage in the tube will increase as the cathode bias is decreased, resulting in higher amplification and greater sensitivity of the circuit to correction signals.

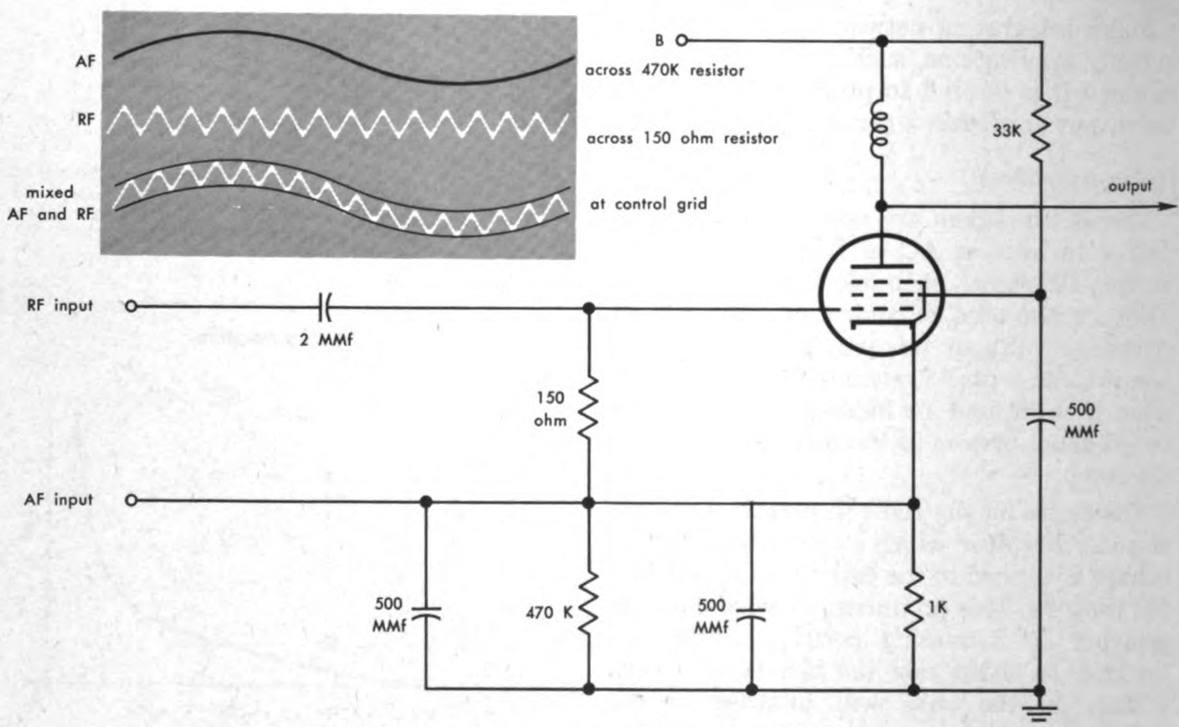


Mixing amplifier





Mixing amplifier as used in a telemetering system



RF-AF mixer and amplifier

Another mixing amplifier, as used in some telemetering systems to combine from two to sixteen frequency modulated audio signals, is shown in the figure on the left.

The 10-K potentiometer acts as the master gain control for the circuit. The mixed signals are amplified in two RC coupled stages employing medium-Mu triodes "V_1" and "V_2."

The cathode follower output stages "V_3" provides a low impedance output. Negative feedback from the cathode follower output stage is applied to the cathode of "V_1" through a 10-K resistor and a .05-mf capacitor in series. This reduces harmonic distortion to a low level.

The frequency response curve of this amplifier is flat within one decibel out of 85 kc.

An amplifier suitable for combining an RF signal with an AF signal is shown in the figure at the left. Amplifiers of this type are commonly found in AM transmitters and in telemetering systems.

In this circuit, the audio frequency signal is developed across the 470-K resistor and the 500-mmf capacitor. The radio frequency signal is developed across the 150-ohm resistor. No RF is seen across the 470-K resistor since the reactance of the 500-mmf capacitor is so low at radio frequencies (10,600 ohms at 30 kc) that it by-passes the RF to ground.

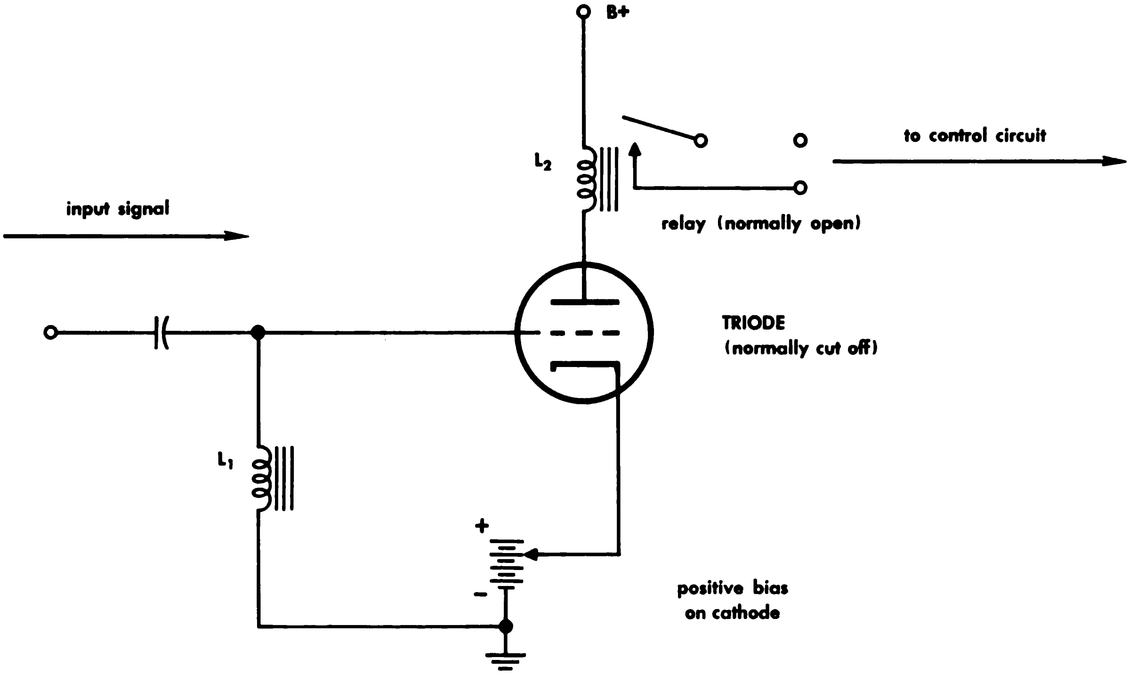
The combined signal is then applied to the grid of the pentode where it is amplified. A resistor (1K) in the cathode circuit shunted by a 500-mmf capacitor provides self-bias. An inductive load is used in the plate circuit to develop the output signal which is taken from the plate.

Component values shown are typical for the circuit as used in telemetering systems. They will vary in accordance with the range of RF and AF signals which are to be mixed.

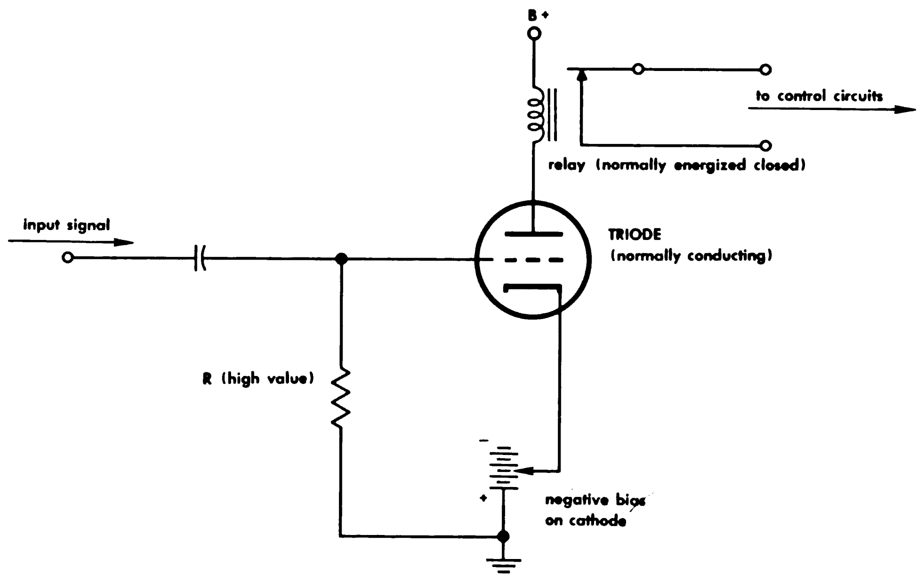
CONTROLLER UNITS

Relays find many applications in missile guidance systems as controller units. They provide a method for utilizing relatively weak radio signals to switch *on* or *off* circuits in which large values of current are present to actuate control mechanisms.

In conjunction with a vacuum-tube circuit, the relay constitutes an electro-mechanical switch which is timed or delayed with respect to the actuating impulse.



Simple relay-closing circuit



Simple relay-opening circuit

Relays are employed in channel selector circuits, arming circuits, and dive circuits. In these circuits, a certain sequence of operation must be initiated by a pulse or signal received from a command transmitter located at a ground station or from an error voltage developed in an automatic navigation system within the missile.

A simple electronic-mechanical plate circuit relay is shown on the preceding page. The circuit, with modifications to suit timing and power requirements, might be employed in any of the previously mentioned applications.

The triode is biased to cutoff by a positive voltage applied to its cathode. When a signal is received, the grid is driven sufficiently positive to overcome the fixed bias, and the tube conducts. The plate current energizes relay solenoid "L_2," causing the contacts to close.

If the input signal is of high frequency, the grid circuit of the tube should have low DC resistance so that when the grid goes positive and draws grid current, it will not develop a bias which would aid the fixed bias and prevent the tube from conducting. A choke (L) is used in the grid circuit to provide high impedance to the signal and low DC resistance to grid current.

The figure above illustrates a circuit in which the relay is normally energized and in which a signal serves to cut off the plate current through the tube. This deenergizes the relay, permitting the contacts to open.

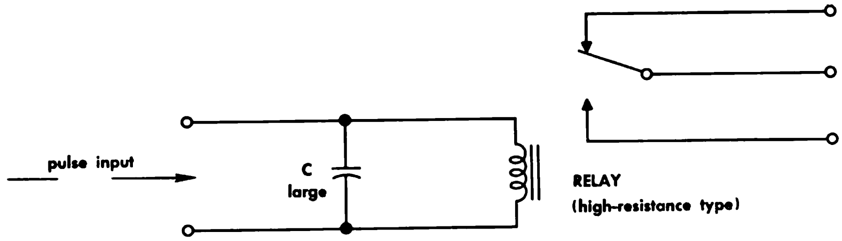
In this circuit, the cathode is biased with a negative voltage adjusted to maintain sufficient plate current to keep the relay closed under *no signal* condition. When a signal is received, the large resistance in the grid circuit develops a high bias, causing the tube to cease conducting. The relay contacts then open.

The figure top right illustrates a relay circuit in which the opening or closing of the relay is delayed. This is done by means of a capacitor shunted across a high resistance relay.

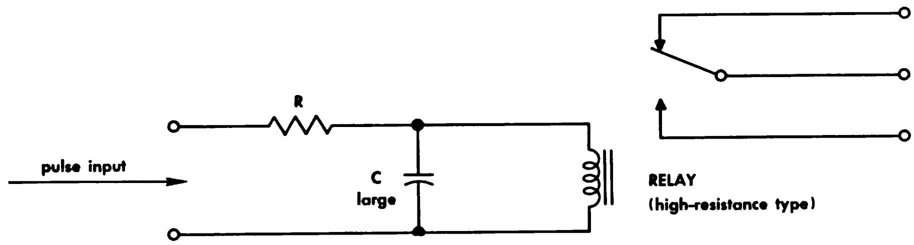
In case a delay in the releasing of the relay is desired, a high-resistance relay is shunted with a large value of capacitance, as shown in the figure.

When an input pulse is received, the relay closes, and at the same time the capacitor takes a charge. At the termination of the pulse, the capacitor discharges through the solenoid coil of the relay and keeps the contacts closed for a period of time. The period of time is determined by the resistance of the relay coil and the value of the capacitor.

Delayed-closing relay

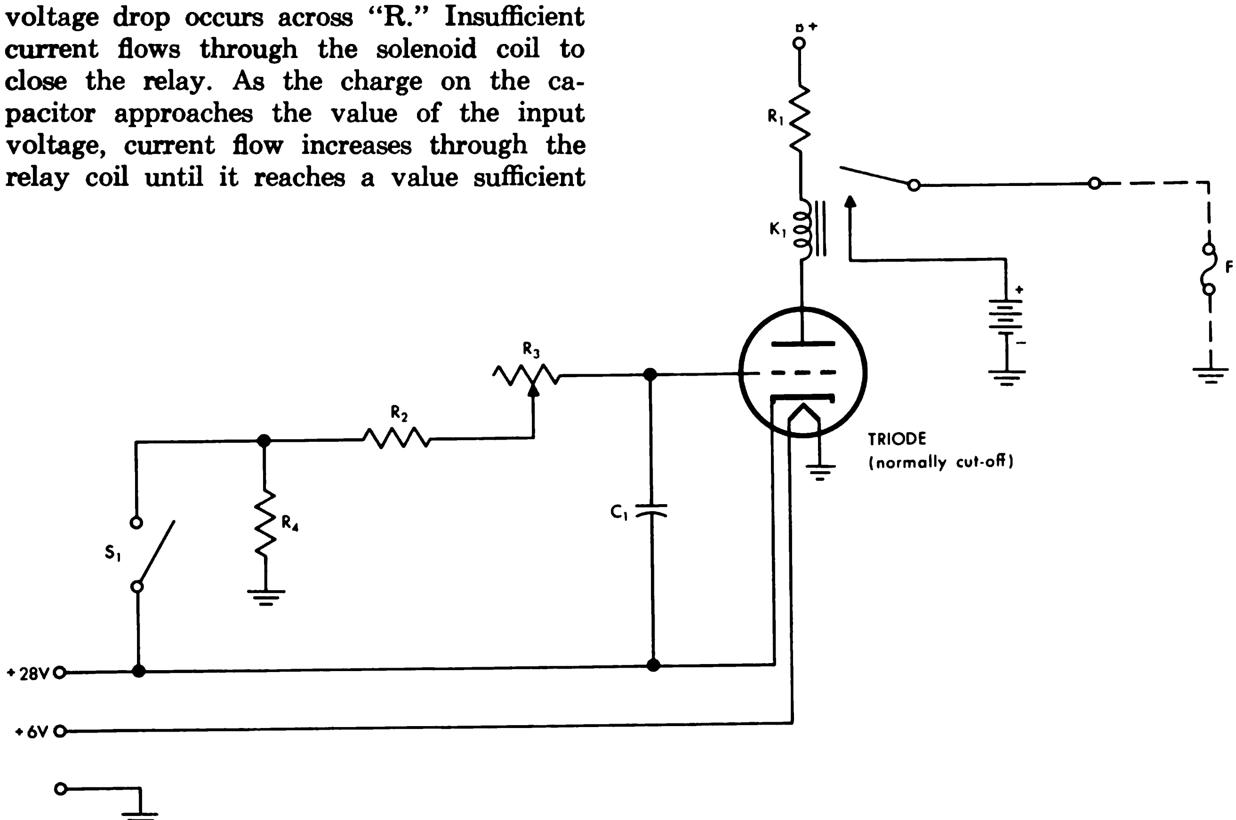


Delayed-release relay

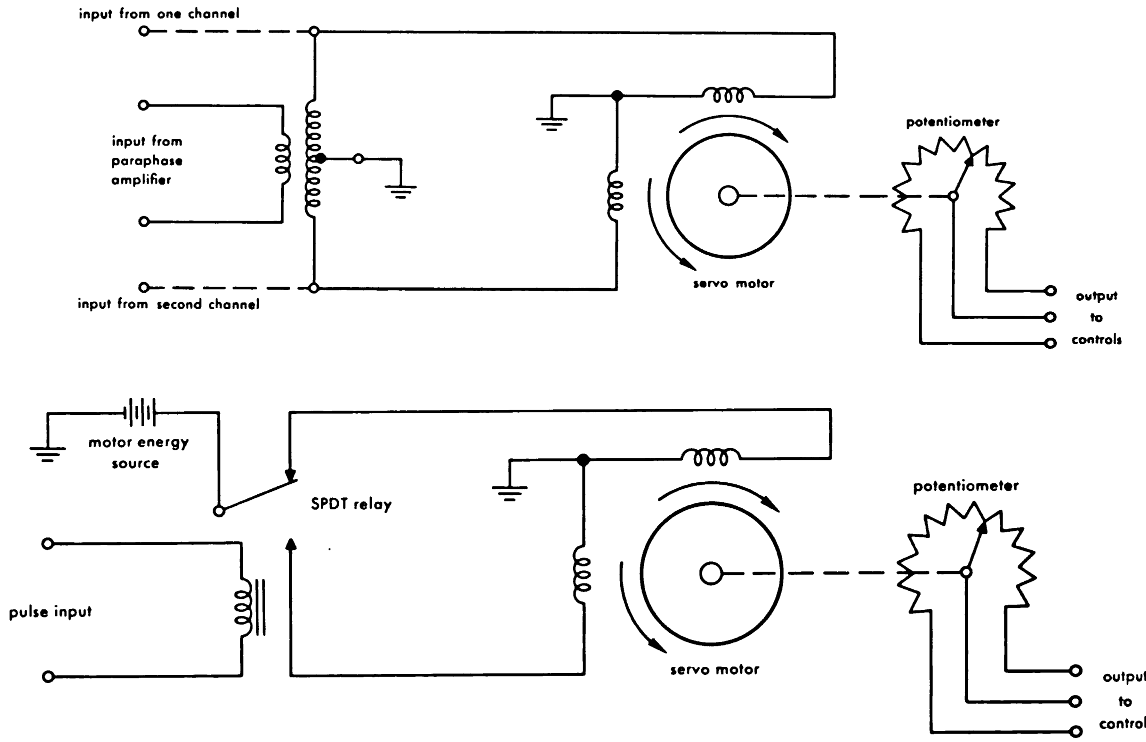


If it is desired to delay the closing of the relay in the preceding example, a resistance is inserted in series with the pulse input as shown above.

When the input pulse is received, a large voltage drop occurs across "R." Insufficient current flows through the solenoid coil to close the relay. As the charge on the capacitor approaches the value of the input voltage, current flow increases through the relay coil until it reaches a value sufficient to close the relay. Thus, the closing of the relay is affected by the charging time of the RC circuit. If the resistance of the relay coil is low, the effect of the RC circuit is less



Time delay-relay circuit



Pulse delay-servo motor activating unit

appreciable; therefore, a high-resistance relay should be used in this application.

Two of these delay-type relays (delay-open and delay-close) are used in parallel for decoding a sequence of pulses.

A relay circuit suitable for application as an arming circuit or gyro-uncaging circuit is shown on the preceding page.

This circuit consists of an inertially operated set-back switch (S) and a triode tube (V). It consists of a normally open relay for applying power to melt the fusible link (F) in the arming mechanism or gyro cage. And it has an RC network and suitable power sources.

The inertial switch (S) is set to trip at the desired acceleration. This removes the cutoff bias from triode "V" after a time determined by the values of the RC components. Once the bias is removed, V_1 conducts, closing relay "K" in its plate circuit. This completes the arming or uncaging circuit.

Normally V_1 is biased to cutoff by the positive 28 volts on its cathode. As the inertial switch closes, the cathode is connected to the grid through R_2 and R_3, rather than through the 28-volt power supply with R_2, R_3, and R_4 removing the bias.

At the same time, R_4 still remains in the circuit between cathode and ground, providing self-bias which prevents excessive plate-current flow through "V."

A time delay occurs between the closing of "S" and conduction of the tube. This delay is determined by the charging time of capacitor "C" which must charge to a specific voltage before the tube can conduct.

The amount of time delay is determined by the adjustment of a variable resistor (R3) which sets the time constant of the RC combination within the limits of its range of adjustment.

ACTUATOR UNITS

Servo motors are utilized in missile guidance systems for operating controls such as variable capacitors, inductors, or potentiometers. These controls correct outputs of range, ground speed, and azimuth circuits. The servo system is designed to provide motor rotation in response to error signals from any two

channels. The motor then drives the potentiometer or other correction device until outputs are balanced, or until a NULL point is reached. When a null point is reached, the motor ceases to rotate and controls remain constant until another error signal is received. A simple circuit illustrating this application of a servo motor is shown on the left.

A servo motor is a reversible, adjustable speed motor. It is controlled in speed and direction of rotation by the phase and magnitude of an *error voltage* produced by variation or unbalance between the outputs of two system channels. When the missile is following the prescribed flight pattern, the outputs are equal in amplitude and 180 degrees out of phase, thus cancelling each other or producing a null output. Any variation in the output of either channel produces an output signal or error voltage which causes the motor to rotate in the direction determined by the polarity of the error voltage. The motor in turn moves a potentiometer which regulates the amount of correction voltage applied to the associated control circuit until a null or balanced condition is restored.

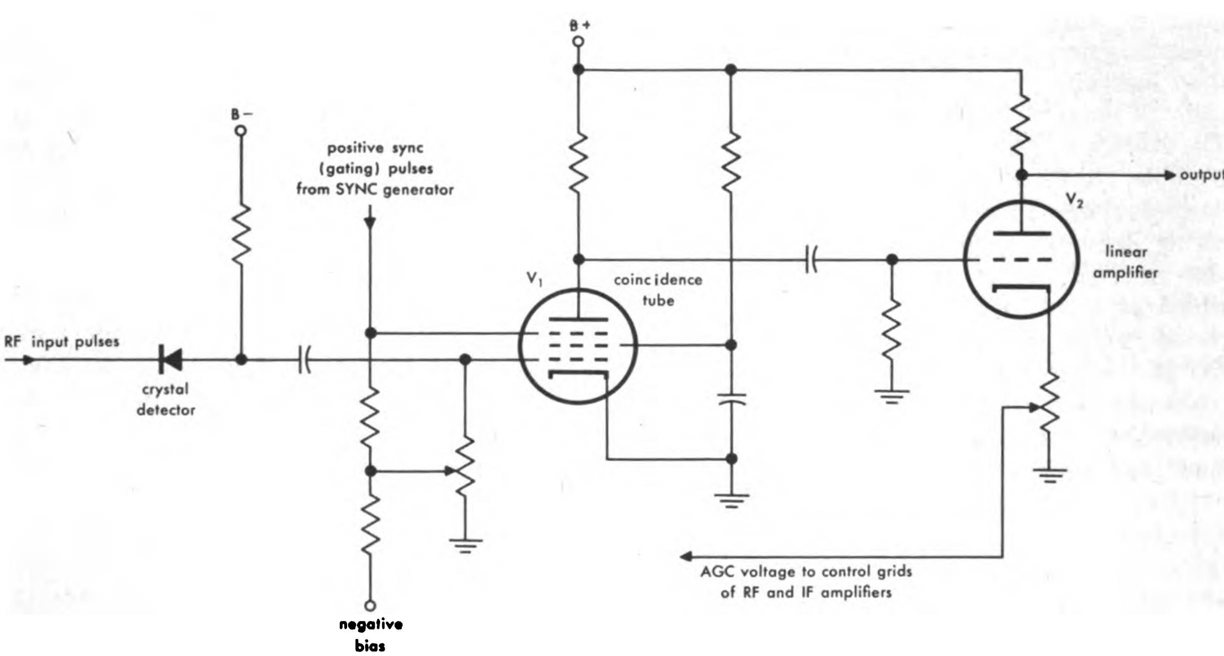
An analogous circuit employing a relay, and suitable for application with circuits having a pulse-form output, is illustrated on the left.

When narrow pulses are received, the relay closes for short intervals on the *make* contact and, because of the long spacing between pulses, closes for longer time intervals on the *break* contact. If each contact were connected to the windings of a motor so that the *make* contact caused rotation in one direction (forward) and the *break* contact caused rotation in the reverse direction, the motor would move in the reverse direction with the first command pulse. With the second command pulse, the motor would move in the forward direction. If the pulses and spacings were equal, the motor would not move.

This circuit is used to decode certain types of command signals and to actuate compensating devices or controls.

FEEDBACK UNIT

Automatic gain control (AGC) is employed in superheterodyne receivers to maintain con-



Gating circuit for applying AGC to RF stages for duration of signal only

stant amplification in the RF and IF stages, regardless of variations in strength of the received signal. This is accomplished by maintaining the correct bias voltages on each controlled stage.

In case the received signal is weak, it is desirable to apply no bias to the RF stages. This permits the stages to operate at their maximum degree of sensitivity. However, if the signal is strong, it tends to overdrive the amplifiers producing distortion and excessively high levels of noise to signal in the output.

If the RF stages were operated at maximum sensitivity at all times, the ratio of noise to signal might be so high that the signal would be lost. The AGC system is designed to provide the highest possible signal to noise ratio over the range of operating frequencies.

Deficiencies of most simple AGC systems are their inability to react quickly to rapid fluctuations in signal strength and their lack of stability when noise pulses are present. These deficiences are particularly noticeable when the signal is in the form of pulses, or is pulse-modulated.

In pulse transmission, the changes in signal strength are rapid, and the duration of the pulse is short. The receiver is subject to noise impulses for a long time compared to the duration of the signal; therefore, the ratio of noise to signal in its output is high. To overcome these deficiencies as much as possible, *keyed* or *gated* AGC systems have been devised. In these systems, a synchronizing pulse is applied simultaneously with the received signal to a coincidence tube which *gates* or energizes the circuit only for the duration of the signal pulse. This excludes the greater portion of the noise, and when the AGC voltage is developed, it is developed to a greater extent by the signal itself and to a lesser degree by noise. The AGC circuit can be designed so that no bias voltage is applied to the RF amplifiers until the signal reaches a certain level then the bias will be applied only after the output has reached the desired level.

There are many methods by which AGC voltage is developed and applied, but only one method suitable for use in radar receivers is discussed here.

The accompanying circuit on the preceding page illustrates a gating system in which the RF signal is detected in a crystal detector and then applied to the control grid of a coincidence (gating) tube. This tube has sufficient negative bias applied to both its control grid and suppressor grid to prevent the tube from conducting due to the presence of a positive pulse on either grid alone. But the tube will conduct whenever positive pulses appear simultaneously on both grids. The tube conducts for the duration of a signal if a positive synchronizing pulse, generated at the same PRF as that of the radar transmitter, is applied to the suppressor grid of the coincidence tube (V_1) at the same instant that the signal pulse appears at the control grid. The signal thus *gated* is applied to the control grid of an amplifier (V_2). As the cathode current varies with the signal voltage, the voltage across the cathode resistor rises and falls proportionally to the variations of signal amplitude. Suitable AGC voltage is picked off by tapping this resistor so that the desired amount of AGC voltage is obtained by any signal level within the limitations of the circuit. By using potentiometers to set the bias level of the coincidence tube and to regulate the cathode voltage of the amplifier stage, the amount of AGC voltage applied to the RF stages is selected. This produces the desired degree of sensitivity in the RF stages when the signal is present and permits a condition of maximum sensitivity between signal pulses.

THE DISCUSSIONS AHEAD

With this brief coverage of amplifier, controller, actuator, and feedback units, we have rounded out our considerations of missile guidance system components. We are now ready to consider the overall make-up of missile systems.

To this point, the text has emphasized the individual functions of control and guidance system components. The next chapter begins a study of guided missile control systems. And then, after covering several types of control systems, the manual introduces you to missile guidance systems.

Chapter 7

Guided Missile Control Systems

This chapter is devoted to covering different types of control systems. The systems discussed are typical of those in the missile field; *no attempt* is made to cover every missile control system. Such an attempt would result in excessive repetition since all control systems are similar in many ways. Furthermore it would confuse you with excessive details.

It is hoped, instead, that in studying this section, you will be able to take any present or future control system and closely relate it to some system in this manual with which you have familiarized yourself. This will enable you to analyze the operation of a new system with a minimum of time and effort.

A knowledge of the operation of a system with which you are involved will provide background enough to start you on the road to understanding not only how a certain procedure is performed but *why*. Such knowledge will produce an increased confidence and interest in your missile work and will enable you to learn other jobs more rapidly. It will enable you to perform jobs beyond normal duty, if such duty should be required due to emergencies.

Numerous operations on missile equipment are necessary from the time it is fired. Assembly, test, and firing procedures for the control, propulsion, guidance, and launching systems must be highly coordinated. Knowledge of system operation will enable you to understand what part some job, though possibly routine, plays in the firing of a missile.

In this manual, the four common types of control systems — *pneumatic*, *pneumatic-electric*, *hydraulic-electric* (or electrohydraulic), and *electric* — are covered. For more detailed understanding of the operation of a system, you are referred to the earlier chapters where individual components of a control system are explained.

WHAT IS A CONTROL SYSTEM?

In the case of a missile, a control system consists of those components which control the missile in such a way as to maintain the selected heading in flight. Most manufacturers and using agencies distinguish between a *control* and a *guidance* system. It is sometimes difficult to determine which portions of the missile system belong to guidance and which belong to control. The difficulty arises because either the dividing line is not definite or some components could be considered a part of both the control and guidance systems.

Generally, a control system is concerned with providing smooth *stable flight*, but it does not know where to direct the missile. The guidance system provides the *location information* necessary to direct a missile to the desired destination. It is necessary for the guidance system to operate *through* the control system in directing the missile to the proper destination. This results in close tie-ins between the two systems. In many cases the same control system is operated with different types of guidance systems and vice-versa. This fact makes a division between control and guidance logical for most types of missiles. Different manufacturers and engineers have various names for a missile control system such as *servo system*, *automatic pilot*, *flight-stability system*, and *attitude-control system*.

REQUIREMENTS OF A CONTROL SYSTEM

The job that a missile control system has to do is not an easy or simple one. The task was mentioned in the introduction to Chapter 5. The control job is to maintain a high-speed missile in stable and smooth flight regardless of outside disturbances; in piloted aircraft this job is normally performed by a human pilot. The problems of stable flight are mentioned in the earlier chapter on aerodynamics.

One of the largest single projects in research and development laboratories during this technical age is that of duplicating the activities of humans for control (or animals for power) by means of machines. The increased use of automatic machine tools in factories is an example of the duplication of human activities by machines. The tractor has replaced the old plow horse. The automobile has replaced the horse and buggy. Automatic control systems are not only common to missiles, but also are found extensively in industry and many mechanical devices.

Many human activities are perhaps more complicated than we are aware of. Furthermore, little is known today, even by specialists, as to exactly what happens in the human mind and body when engaged in activities such as picking up a piece of paper, driving a car, or flying an aircraft. Experts agree that the mental and physical processes are highly complex. Therefore, duplication of certain functions by means of automatic systems is not simple.

In driving a car, for example, if a turn must be made, you *remember* how you made a turn before. You think automatically "If I made a similar turn by a certain method before, I can use that same method again." You then proceed to determine how sharp to turn, depending on the situation compared again to previous situations. The amount of turning is therefore *anticipated*. As you turn, you may alter your degree of turn depending on developments. This provides almost *immediate* control of the wheel. As you complete the turn, you decide how soon and how fast to return the steering wheel back to normal. Your speed, weight of the car, sharpness of the curve, and condition of the road, are all factors to be considered. At the same time you are thinking whether any braking, speeding up, or shifting of gears is necessary. You can see, then, that a control system designed to automatically steer a car around a curve would not be simple. Certain equipment can be compared to the performance during driving. In a sense, remembering is done by *integrating devices* and anticipation by *rate devices*. Application of these devices is covered later.

Flying an aircraft is a complex human activity, and it certainly requires much design technique to create a system which will replace a human under all flight conditions. *Such a system must be rigorously checked out before it can be depended on to control a missile.* The last section of this chapter covers basic checkout procedures. Fortunately for the missile development program, systems were designed which flew aircraft automatically before the

need for a similar system in missiles existed. These *automatic pilots* were installed in aircraft to relieve the pilot of excessive fatigue caused by long flights, thus increasing his efficiency when peak performance was required. Actually, automatic pilots have performed better than human pilots under certain conditions, but they cannot exercise judgment in emergencies. Autopilots were adapted for use as control systems in some early missiles. What remained to be done was the development of lighter, faster responding systems, which could accommodate the higher speeds and accelerations of missiles.

It is not difficult to analyze what a control system must do. First, it must have some means of sensing or discovering when something is wrong with the object to be controlled. This object to be controlled is sometimes called the *plant* or *load*. In the case of missiles, the object to be controlled would be the craft; it would be controlled in regard to its attitude, altitude, and speed.

Second, the system must be able to distinguish in which direction from the desired point the error is occurring and to correct accordingly. For example, a correction for right deviation certainly cannot be the same as for a left deviation. This is known as the sense of the error. In an electrical system, sense is usually distinguished by using voltages of either opposite polarity or opposite phase. This detected error is represented by a signal that usually is a small electrical voltage and is known as an error signal or as information.

After a ~~missile~~ error "symptom" has been relayed to the controls, a "cure" must be performed. The control system must produce action which will properly remedy the error. This means it must operate the control surfaces or similar controlling devices. To operate the devices properly, it may be necessary to change the form of the error signal so it represents additional information. Next, this error information signal must be amplified to a signal which has enough power to operate the control devices, such as control surfaces. The power is transferred by high-pressure pneumatic lines, hydraulic lines, or by wires.

The control system must produce corrective action, without too much delay from the time the error began. This period of delay is known as the *speed of response* of a system.

After the surfaces begin to move, the control system still has not completed the job. During the entire period of missile correction, the system is continually detecting the instantaneous missile errors and making constant adjustments of the control surfaces. Some systems also detect the speed that the missile is deviating from the desired heading; they take this factor into consideration when positioning the control surfaces. This constant vigilance indicates that a well designed system could do a better mechanical job of flying than that produced by the slower reflexes of a human being.

TYPES OF CONTROL SYSTEMS

Four methods of aerodynamic control are: *control surfaces*, *jet vanes*, *movable jets*, and *fixed steering jets*. These methods are discussed in chapter 2.

In the case of control surfaces, jet vanes, and movable jets, the control system would deflect these devices to change the forces on the missile. When using fixed steering jets, the missile is controlled by varying the thrust from each jet. The rate or time of burning would be controlled by valves operated by the control system.

Control systems could possibly be classified into the four categories above. However, this would not be the best division. The equipment required to move the control surfaces, jet vanes, and movable jets may be similar or different, regardless of whether jet vanes or control surfaces, etc., are used. Actually, the major difference in control systems is in the *method of moving* these attitude control devices.

There are three methods of producing this motion: (1) by an *air* (pneumatic) piston, (2) by a *hydraulic* piston, or (3) by an *electric* motor. The type of control system, then, is determined by whichever method, or combination of methods, is used to produce motion in the system. *This is the most logical division because the checkout of the system and the type of components used* in the system depend on which method, or combination of methods, is used.

The selection of the control system for a particular craft depends on factors such as the speed, size, altitude, range, and weight of the missile. Incidental factors, such as the professional preference of personnel of the engineering department designing the missile, also influence selection. If they are more familiar with certain components, for example, they may be more inclined to apply these components to the system. Possibly, a company may already possess equipment relative to a particular type of control system. In other instances, the USAF contract for development of a missile will specify the type of control system.

Some of the advantages and disadvantages of particular types of systems are quite prominent. A pneumatic system which depends on tanks of compressed air is obviously limited in range. Since air or any other gas is compressible, the movement of an actuator is slow due to the time it takes to compress the air in the actuator to a pressure sufficient to move it. Hydraulic fluid is practically incompressible and will produce a faster reaction on an actuator, especially when the actuator must move against large forces. This asset is evidenced by the fact that large, high-speed missiles are controlled by hydraulic actuators.

A hydraulic system normally weighs more due to the need for a pump, a reservoir, filters, and an accumulator. Also a hydraulic system is hard to maintain, requiring filling and bleeding operations. Bleeding is done to remove air bubbles from the hydraulic fluid.

At the high altitudes at which some missiles are intended to fly, new problems of control-system design are created. Temperature and pressure are severely reduced. This has an effect on any type of control system. Hydraulic fluid may thicken to the point of being inoperative. Air bubbles in metal parts may expand and create malfunctions. High-altitude lubrication of mechanically moving parts must be considered. Changes of temperatures also affect the operation of electronic parts. Altitude must, therefore, be considered in selecting the best type of control system for a particular type of missile.

The use of an all-electric system has important advantages. Very few missiles have been designed, or will be designed, which do not have some part which operates by electricity. The use of an all-electric control system would place all the missile equipment, except the propulsion unit, within the electrical field. This would simplify manufacture, assembly, and maintenance. Also, it would be easier to transmit information or power to all parts of the missile by wires, rather than hydraulic or pneumatic tubing.

A mechanical control system in a missile is not very probable. Consider first what would constitute an all-mechanical system. Error information would be transferred from a mechanical sensor by some mechanical means such as a gear train, cable, rotating or sliding shaft, or chain linkage. This linkage would then connect to the correcting devices such as control surfaces or movable jets. Any computing devices in the system such as mixers and integrators would have to be mechanical.

There are four major disadvantages of such a system. The first disadvantage is that *insufficient power* is applied using all-mechanical linkage. A mechanical advantage could only be gained by means of levers or gears. It would, therefore, be hard to obtain enough power from a reference unit to produce sufficient movement of control surfaces.

A second disadvantage results from the *difficulty of installing these linkages* from the reference unit through the system to the control surfaces of the missile. For example, it is much easier to install electric wires or tubes throughout a missile than to install moving cable or shafts.

Furthermore, gears, shafts, and cables strong enough to exert force on control surfaces *weigh* more than wires or tubes which are capable of transmitting the same amount of power. Weight is a serious consideration in a missile. The addition of a few extra pounds will "snowball" into more added weight by the resulting necessity of adding more fuel and possibly a larger propulsion system.

A fourth disadvantage is that it is not possible to *pickoff error signals* from a reference unit mechanically without a disturbing force imparted to the unit. Most reference units — whether they be gyros, altimeters, transducers, or compasses — are sensitive

and would be inaccurate if a mechanical coupling (pickoff) were used.

However, in the new dynamic field of missiles little can be called impossible. Many past dreams have become realities. Also, many ideas are leaving the realm of fantasy to enter the realm of possibility.

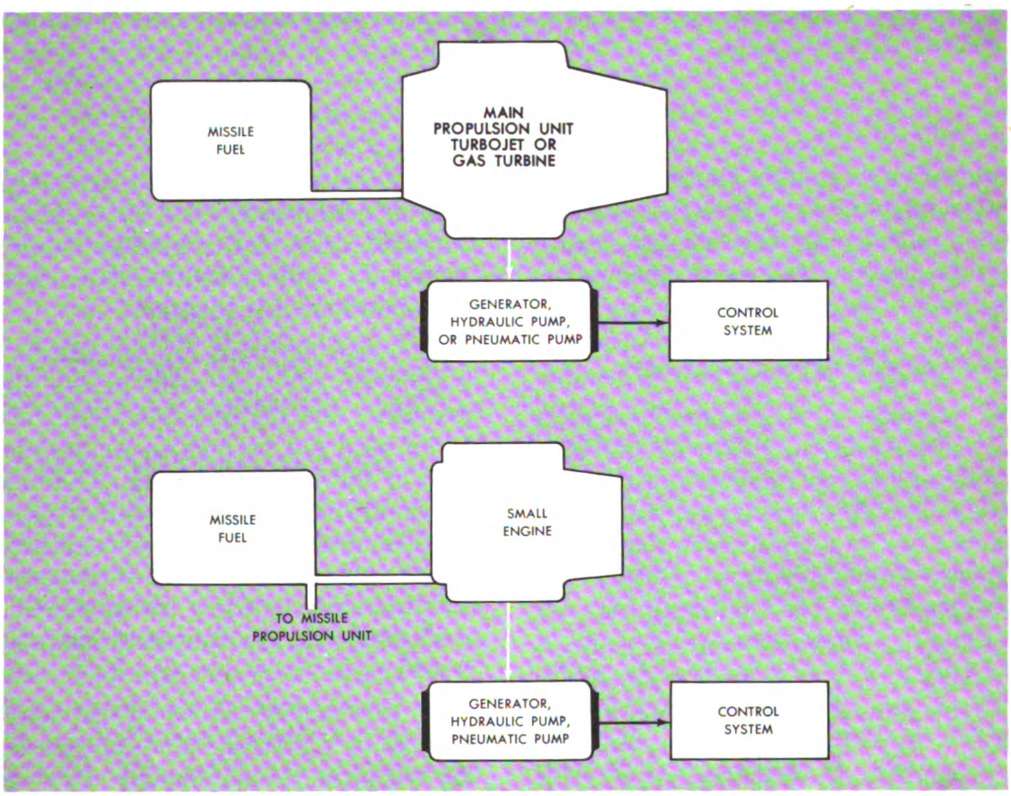
All control systems are vulnerable to enemy defensive fire. The best protection against defensive fire consists of high altitude, high speed, and a large salvo of missiles. If any part of the high-pressure portion of a hydraulic or pneumatic system were damaged, the complete control system would probably become inoperative. This condition would occur if a bullet should strike the reservoir, lines, or almost any of the components.

CONTROL-SYSTEM ENERGY SOURCES

Energy is required to operate control surfaces, jet vanes, movable jets, and fixed steering jets. This energy must come from a source within the missile, since no system has ever been discovered to efficiently transmit large amounts of power through space without a guide such as wires, tubing, etc. Regardless of the type of control system, the energy to operate the system must come from one of three sources.

Energy from Missile Engine

As one possibility, the energy may come from the propellent as shown in the accompanying figure. This is the most common energy source. If a turbojet or gas turbine engine is used to drive the missile, a generator is mechanically connected to the turbine to provide electrical power. A generator may also be used as a starter. Similarly, a hydraulic pump or air compressor is connected to the turbine to provide hydraulic pressure or air pressure. This system is practical despite the fact that the accessories deprive the thrust source of a little power. Control power, like thrust power, is derived from the energy of the missile fuel. This system produces a comparatively large amount of



Two methods of obtaining power from energy of fuel

power for the control system, considering the weight of fuel used.

Energy from Auxiliary Engine

The fuel may also be used directly to drive a small auxiliary engine which drives a generator, hydraulic pump, pneumatic pump, or combination of the three units.

Stored Energy

A possibility for electrical power is to use batteries. This is applicable to short-flight missiles.

Another source of stored energy could be provided for by using tanks of compressed gases; this source is also more applicable to missiles of relatively short flight. These compressed gases are used to actuate the controls directly by air connections through control valves to the actuators. The gases may also provide power *indirectly* by operating a small turbine or maintaining hydraulic pressure by means of a pressurized accumulator.

Now you are ready to consider specific types of systems. The following section covers pneumatic and pneumatic-electric control systems.

Chapter 7 • Section 1

Pneumatic and Pneumatic-electric Control Systems

Even though pneumatic control systems are not commonly used in missiles today, it will be helpful to look into this system before taking up the more widely used pneumatic-electric systems.

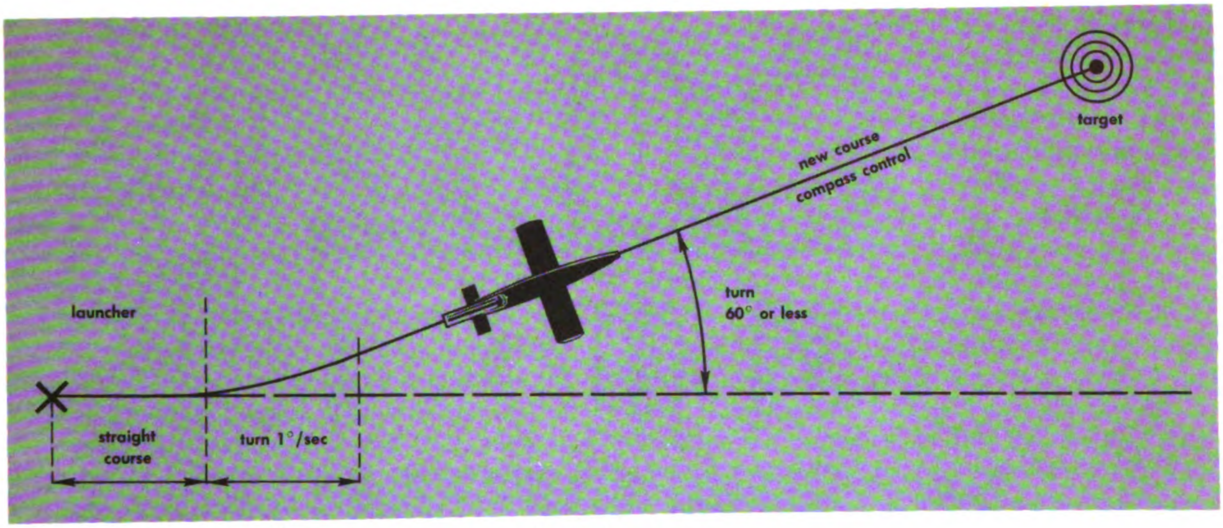
PNEUMATIC CONTROL SYSTEMS

A good example of a pneumatic control system is that used in the USAF JB-2, which is used both as an experimental missile and target drone. The JB-2 was developed during the beginning of the period when the missile field developed rapidly. Even so, it represents many good pneumatic control principles.

The JB-2 is launched from a ramp at an angle of about 7°, using a booster assembly. After it leaves the launcher it begins a slow climb until it reaches the altitude preset by barometric control. It then maintains level flight until an air-log device produces dive or dump signals which extend elevator spoilers to dive the missile into the target. A preset system of guidance is used by making altitude, direction, and distance adjustments prior to launch. A typical flight pattern is shown in the accompanying diagram.

Provision is made for the missile to make a right or left turn early in the flight. This provision has many advantages. The launch ramp need not be pointed in the exact direction of the desired heading. This means that the launch ramp can be rather permanently fixed and missile can be launched at different targets from the same launcher. Furthermore, heading to a target changes with time due to changes in prevailing crosswinds.

A mechanical timer is adjusted before launch to determine the time at which the turn begins. It can be any time up to five minutes after launching. Another preset timer controls the length of time the missile turns.



Flight plan of JB-2

The length of time for turning can be up to one minute. Since the rate of turn is one degree per second, the maximum turn is approximately 60°. The turn is accomplished by slaving coils which magnetically exert a force on the displacement gyro, thus precessing it in azimuth.

After the turn is completed, the yaw gyro is slaved to a certain heading by a compass which has been preset. This condition continues until dump point at which point the rudder is locked in a streamline position. The displacement gyro is uncaged at the instant of launch. The rigidity of the gyro is a sufficiently stable reference to maintain proper yaw attitude until after the turn. After the turn, the gyro is slaved by the compass. The amount of turn need not be highly accurate since the final heading is determined by the compass setting.

General Operation of a Pneumatic System

Knowledge of the performance and specifications of the JB-2 should indicate that a pneumatic control system is suitable for the missile. The range is short enough to carry compressed air for the entire flight. The missile is relatively small and flies at subsonic speed; therefore, it does not require very much force to move the control surfaces. The time lag created by the cushioning of air is not great enough to rule out the use of air pistons to move the control surfaces.

Thus, the control system is almost entirely pneumatic. The rotors of the gyros are powered by air. The gyros are powered by air. The gyro error pickoffs are all air blocks; therefore, the control information is in the form of varying air pressures. And the control surfaces are moved by air pistons. Air for the entire flight is obtained from two tanks filled to a pressure of 2000 pounds per square inch.

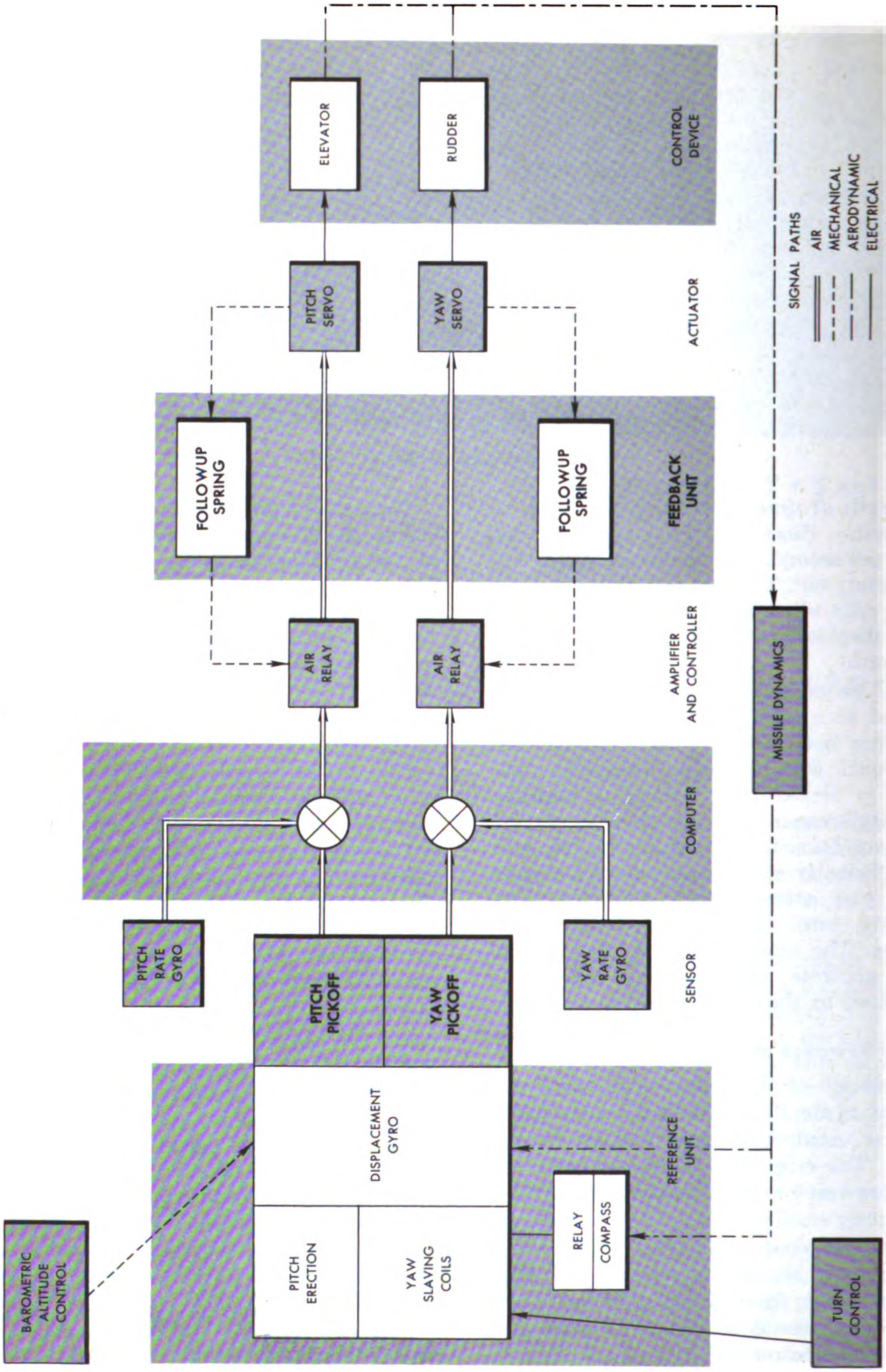
The only error signal which is not pneumatic is an electrical signal which operates the azimuth electromagnets to slave the gyro.

There are advantages in a system which uses only one medium for operation — i.e., all air, all electric, or all hydraulic. When only one source of power exists, equipment such as transducers, transfer valves, or air switches are not necessary to convert from one medium to another.

The JB-2 is controlled by only two sets of control surfaces. *Pitch control* is achieved by the *elevators; yaw* and *roll* are controlled by the *rudder*.

The next diagram, Operation of Pneumatic Control System, shows that the entire system contains three gyros for reference: one displacement gyro and two rate gyros. The diagram should not be memorized, but it can be used as a means for learning system operations.

Starting with the major reference unit, which is the displacement gyro, the pitch and



Operation of pneumatic control system

azimuth missile errors are sensed by air-block pickoffs. It is explained later how the yaw pickoff also detects roll deviation. Each signal is sent by means of varying air pressures to an air relay. The *air relay* acts like a combination *amplifier* and *controller*. The input to the relay is in the form of small changes of air pressure from the air block pickoffs. The output of the relay is a pressure which is high enough to actuate an air piston (servo). The two rate gyros also produce a pneumatic signal which joins with the displacement signal of the respective channel. The creation of the rate signals, and the addition of these signals in the proper ratio with the displacement signals, can be considered to be the *computer functions* of the system.

Any difference between the actual and desired angular position (or attitude) or the missile in relation to the displacement gyro produces the required elevator or rudder movement. This action produces a movement of the missile which in turn affects the relationship of the reference units and the missile. This fact is shown by the block, Missile Dynamics, at the left, and the dotted connecting line which completes the servo loop.

The *followup signal* is actually a mechanical *feedback* force exerted by a spring. Both the diaphragm and spring exert force on the servo valve. Movement of the servo to either side of normal produces a force back on the air relay valve. This action tends to return the servo to the normal or midposition to produce streamlined control surfaces.

The corrective signal to the servo piston must be somewhat dependent on the instantaneous position of the control surface. This position is indicated by the followup signal. The subject of followup and feedback is covered earlier in Chapter 5. Again, a computer function is performed, as spring-force information is combined in proper sense and ratio with air-pressure information at the air relay.

Yaw Control

The pictorial diagram of the pneumatic control system on the following page shows actual air connections to the various units. From this diagram the interaction of the rate and displacement signals can be determined.

Remember that the rate signal is proportional to the *speed* at which a missile deviation is *changing* in magnitude. This is different from the displacement or proportional signal, which is proportional to the missile angular deviation at any instant. At the azimuth rate gyro, the rate signal appears as an unbalanced air pressure between two holes in an air-block pickoff.

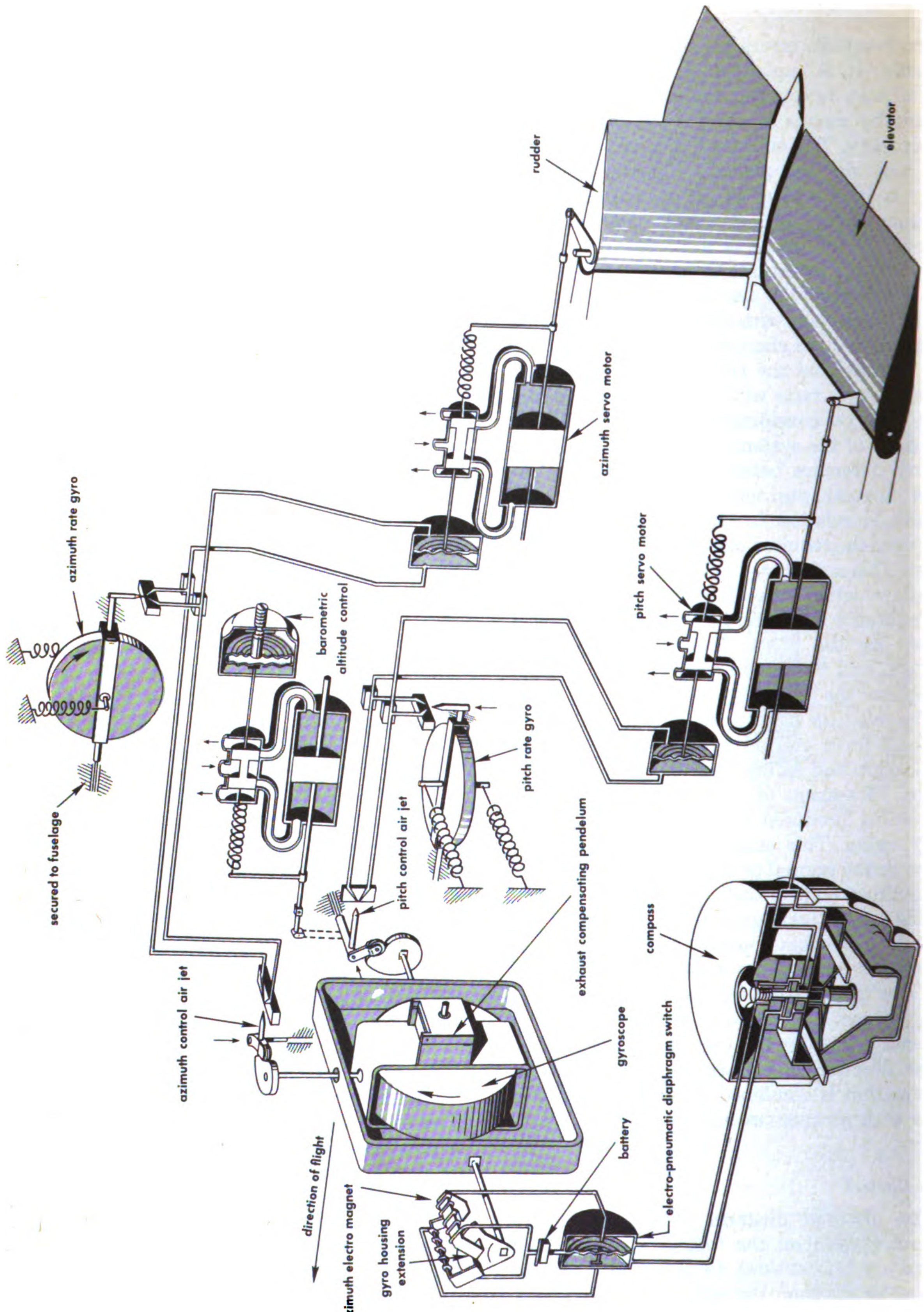
Now consider yaw. Suppose the missile nose veers to the right. A displacement gyro signal develops at the pickoff (azimuth control air jet). The azimuth control air jet pivots to increase air pressure in the left hole of the pickoff (when facing the direction of flight). This air pressure is transported, in the lower of the two air tubes, to the diaphragm of the air relay. The diaphragm is forced to the left. This controls high-pressure air which forces the azimuth servo motor (actuator) to the right. Mechanical linkage moves the rudder to the left, correcting a nose-right deviation.

Also, as the nose of the missile moves to the right, a signal is produced by the azimuth rate gyro. By the law of gyro precession, the azimuth rate gyro exerts more force on the right restraining spring because force on the gimbal precesses the gyro a small amount. As it precesses, more air is received by the left hole. This increases the pressure in the same tube that contains the high pressure signal from the displacement gyro. The rate gyro is *supporting* the corrective of the displacement gyro.

As the missile deviates from its desired heading, the rate signal increases the corrective action of the displacement gyro. Therefore, if the deviation from proper heading were increasing at a rapid rate, the corrective signal from the rate gyro would be large. This reduces the total deviation of the missile.

Consider the instant when the missile has veered as far to the right as it is going to go. The error signal from the displacement gyro is greatest since the error is greatest. There is *no signal* from the rate gyro because the missile is not *changing* its heading at the instant under consideration. The rate gyro has been returned to its mid-position by the restraining springs.

An instant later the error is decreasing as the missile begins to correct. The error from the



Pneumatic control system

displacement gyro still has the same sense; that is, the pressure still is greater in the left hole of the azimuth pickoff because the error is still to the right. The signal is decreasing. However, the error signal from the rate gyro has reversed sense. The missile nose is now moving in the other direction so the force on the rate-gyro gimbal is in the opposite direction. The precessing gyro creates a pressure in the *right* hole of the rate pickoff. The signal *partially counteracts* the displacement gyro signal. The support and counteraction by a rate signal on a displacement signal are illustrated below.

The reduction of the displacement signal by a rate signal leads to two questions. First, "What effect will this reduction of proportional error signal have on the rudder and the missile as the missile is returning to the desired heading?" Second, "Wouldn't it be better to have a rate signal *support* rather than *counteract* the displacement signal at this time?"

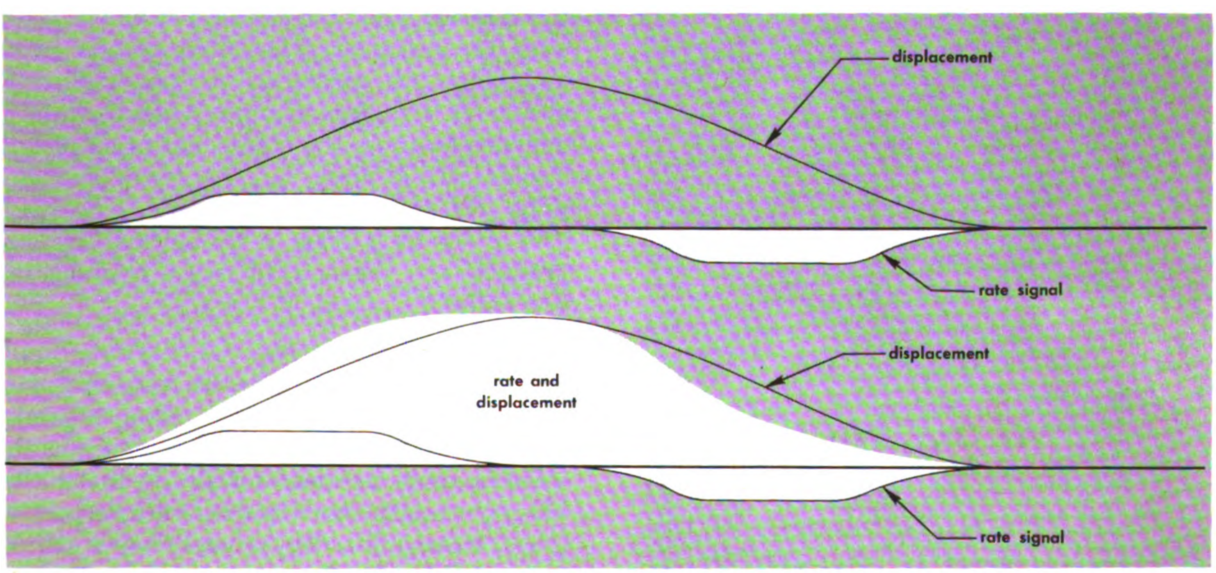
If the signals were supporting, the rudder would be extended further away from streamline, and the missile would be returning to the desired heading at a faster rate. But the missile would be returning so fast that it would swing past the desired heading to the other direction. A veering to the left would then cause corrective action by the control system for *that* direction, and so on and on, wobbling to both sides of the desired heading. Such action is, of course, undesirable and is known as oscillation or hunting about the desired heading. The rate circuit prevents this.

Whether the missile is deviating from the proper heading or returning to proper heading, the rate gyros improve the response of the control system and produce smooth stable flight. When the missile deviates, the rate gyros reduce airfoil correction, thus preventing the missile from returning so fast that it overshoots the desired heading.

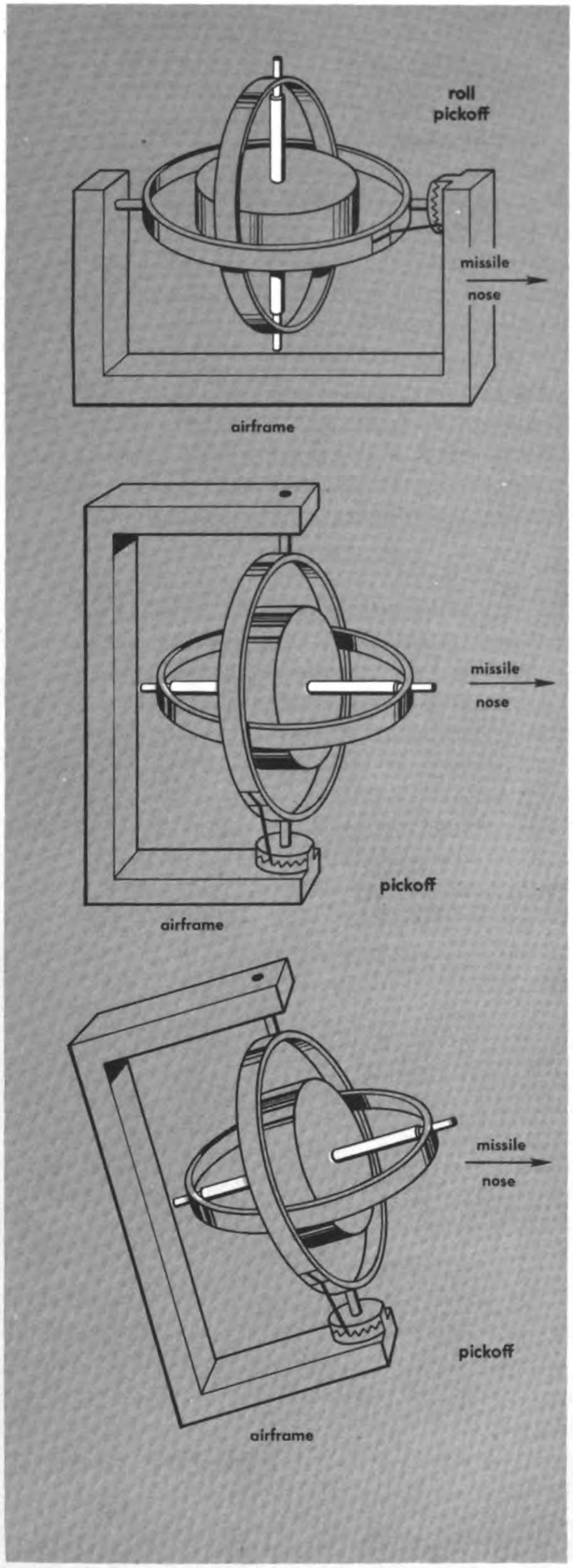
Exactly the same interaction of the displacement and rate signals takes place in the pitch channel. Furthermore, this fundamental action of a rate gyro or rate circuit output is the same for every missile control system.

Roll Control

Attitude errors about all three axes are detected by one gyro. Normally, if a yaw and a pitch pickoff operate from the same gyro, the spin axis is in line with the roll axis of the missile. However, this gyro is tilted backward to an angle of 20° to the horizontal. This angle produces an error in the *yaw* pickoff when the missile *rolls*. The top drawing on the following page shows that the spin axis of a roll gyro would normally be vertical. Note also that the spin axis of a yaw gyro would normally be horizontal.



Effect of combining rate and displacement signals



Obtaining yaw and roll errors from one pickoff

A gyro which is mounted at an angle in the manner of the one in the bottom drawing produces an error due to both roll and yaw. Therefore, the output of the yaw pickoff in this case may contain a component due to yaw and a component due to roll. The ratio of the two errors depends on the attitude of the missile and the angle at which the gyro is mounted. Since the gyro is close to horizontal, a greater signal is produced for a given yaw attitude error than for the same error in roll. However, the roll signal is sufficient to maintain the missile stable in roll by means of rudder action.

Pitch Control

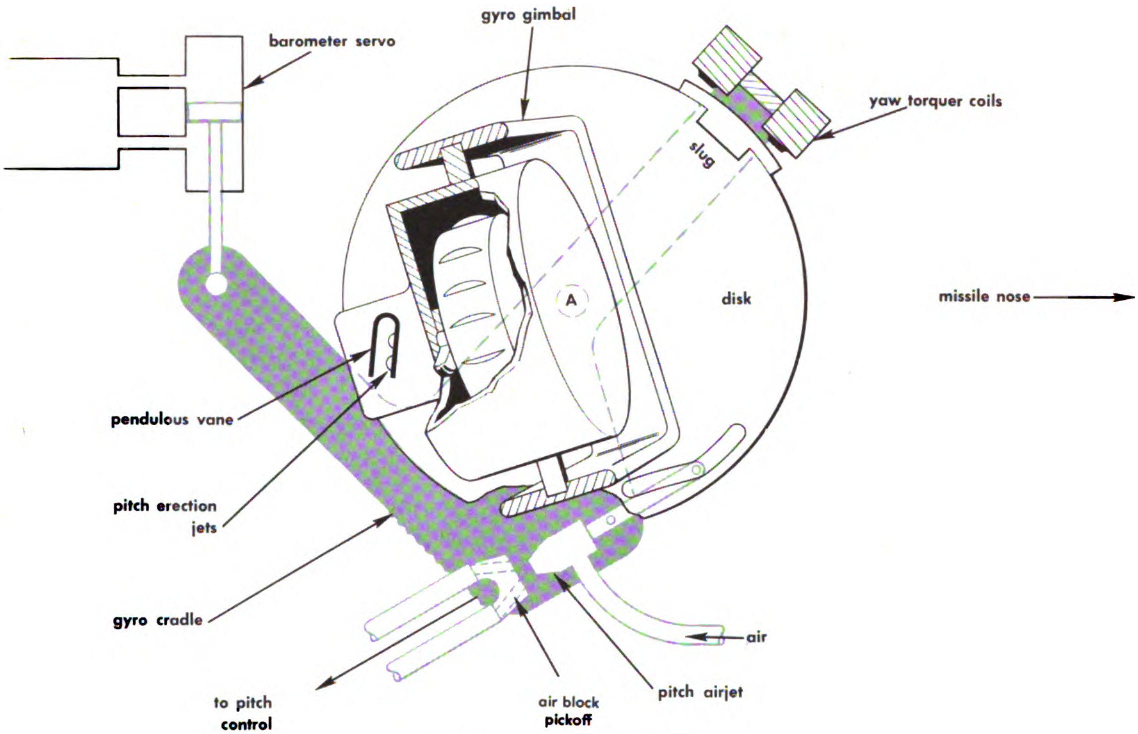
The diagram of the displacement gyro assembly on the right shows the relationship of the pendulous vanes, azimuth yaw torquer coils, pitch pickoff, and barometric altitude control. The diagram shows how the devices operate together. It is, of course, not to scale.

When the JB-2 deviates in pitch, more air is directed into one hole of the pitch pickoff block than the other. This pressure difference represents a pitch error signal which connects to an air relay. The air relay controls air pressure used to move the elevator.

The rotor of the gyro tends to maintain a constant angle in space due to gyro rigidity. This means that the gimbal and gyro disc also maintain a constant angle since they move with the gyro. The disc is rigidly connected to the gimbal. The gyro cradle normally moves with the airframe. The pitch pickoff pivot and block are connected to the cradle and also move with the airframe. When the missile deviates in pitch, the cradle and pitch pickoff also deviate in pitch, but the gyro disc maintains the same position in space. The pitch arm pivots as it rides in the slotted disc and produces the pressure difference between the two holes of the pickoff block.

Barometer Altitude Control

Pitch action can also be caused by a barometer servo that is controlling altitude. As you can see in the preceding diagram, the servo is connected mechanically to the cradle. The gyro cradle remains fixed with respect to the airframe unless the barometer servo should move it. When the barometer servo moves, the position of the gyro cradle with respect to the



Displacement gyro assembly

missile frame changes by pivoting at point A. Also when the gyro cradle moves, the nozzle and block of the pitch pickoff move with it. Since the gimble and disc remain stationary, the pitch pickoff arm pivots and produces a pitch signal. This, of course, also produces elevator movement and pitch reaction.

The stability of the missile in pitch is produced by the relation of the missile and the gyro. The initial climb angle and altitude of the trajectory are controlled by the barometer servo operating the gyro cradle.

A note of caution: Never confuse the control system of a missile with the slaving system for the various gyros. Up to this point, the slaving system has not been gone into in regard to system operation. Actually, it has nothing to do with the control system directly, since it simply corrects for gyro drift. Although the discussion has assumed that the gyro maintained its proper position in space, such is not the case. Therefore, the pendulous-vane erection mechanism, shown in the preceding diagram, is used to keep the displacement gyro at the proper angle with the vertical (the direction of force of gravity).

Gyro Slaving in Yaw

Slaving coils are used to keep the gyro itself on the desired heading. At first glance at the figure above, it appears by the location of the slaving coils (yaw torquer) and slug that they would control the gyro in pitch. This would be true were it not for the fact that a gyro does not act the same as a stationary mass. In the coverage of gyros, the law of precession stated that a gyro will move or precess in a plane that is *perpendicular* to the direction of the applied force. This law is applied when current flows through one of the coils. The current exerts a magnetic force on the slug which creates a force on the gimbal in the pitch plane. This moves the gyro in *yaw.*

The coils are connected to the timing mechanism which energizes them to effect a preset turn. The gyro then precesses in yaw, producing an error at the azimuth air pickoff and giving rudder action.

After the turn is completed, the coils are switched to the magnetic compass slaving system. If the gyro should drift in yaw, relative to the magnetic lines of force of the

earth, one of the slaving coils would be energized until the error of the gyro is corrected. This action occurs as follows:

> Precession of the gyro in yaw causes the missile to follow the gyro and deviate in yaw, causing an error in the compass sensor. Relative movement of the airframe and compass needle actuates an air electric relay which energizes one of the coils.

This movement of the missile brings up a logical question: Is not the compass slaving system acting as yaw control for the missile rather than simply keeping the gyro in the proper heading? An ideal slaving system *should* function entirely separate from missile movement. Most slaving systems do function separately from the control system, though this system does not.

Missile yaw error causes this slaving system to control the rudder since the slaving precesses the gyro. However, the effect is slight compared to the corrective action created by the yaw control system. For example, suppose the missile deviates 6° to the left, due to wind. This actuates the air-electric compass relay, which energizes one of the slaving coils tending to precess the gyro. The direction of precession is opposite to the deviation of the missile, since the slaving system "thinks" that gyro drift is producing the deviation. Before the gyro has a chance to precess any given amount, however, the yaw pickoff has detected the error and corrected the missile.

On the average, the gyro maintains the direction preset in the compass before flight. The missile maintains the heading of the displacement gyro.

Still another question could now arise. Why not use the signals coming from the air-electric compass relay directly as yaw attitude error signals and eliminate a displacement gyro pickoff? First, a pickoff on the compass which provides proportional control (signal proportional to amount of missile deviation) would have to be used. This arrangement is possible. Second, the compass unit would provide no roll information to the rudder which is necessary in this control system. Third, a compass possesses *turning error*. The turning error produces incorrect error signals during the period that the missile is turning. This would not enable a compass alone to be used as the yaw attitude sensor.

You will find that asking yourself, or other people, questions like the previous two will lead to a more complete understanding of the operation of any system which you may be considering.

At this point, we'll leave this discussion of pneumatic systems and take up considerations of control systems that employ electric as well as pneumatic principles.

COMBINING OF AIR AND ELECTRICITY IN CONTROL SYSTEMS

A pneumatic system *need not* operate completely by air. It can be a combination of several types of systems.

An electric-type pickoff is accurate and dependable in providing a signal proportional to displacement. Also, electric information and power can be transported more easily than hydraulic or pneumatic information or power. Much more leeway exists in the use of electrical rather than mechanical, hydraulic, or pneumatic computers. However, it is difficult to design an electric motor which possesses the required speed of response and force to actuate the control surfaces and still be light in weight.

Above factors point to a pneumatic combination system if the medium of air is to be used at all. Electrical components are used in the beginning of the system, and air is used to operate the servos at the end of the system. As already stated, such systems are presently being used in missiles.

The representative pneumatic-electric system, which we will discuss, is suited to a small, subsonic, short-range surface-to-air missile. You will recall that the completely pneumatic system just discussed is suited to a small, subsonic, short-range missile. The actuators of both systems are operated by air.

Actuator Booster

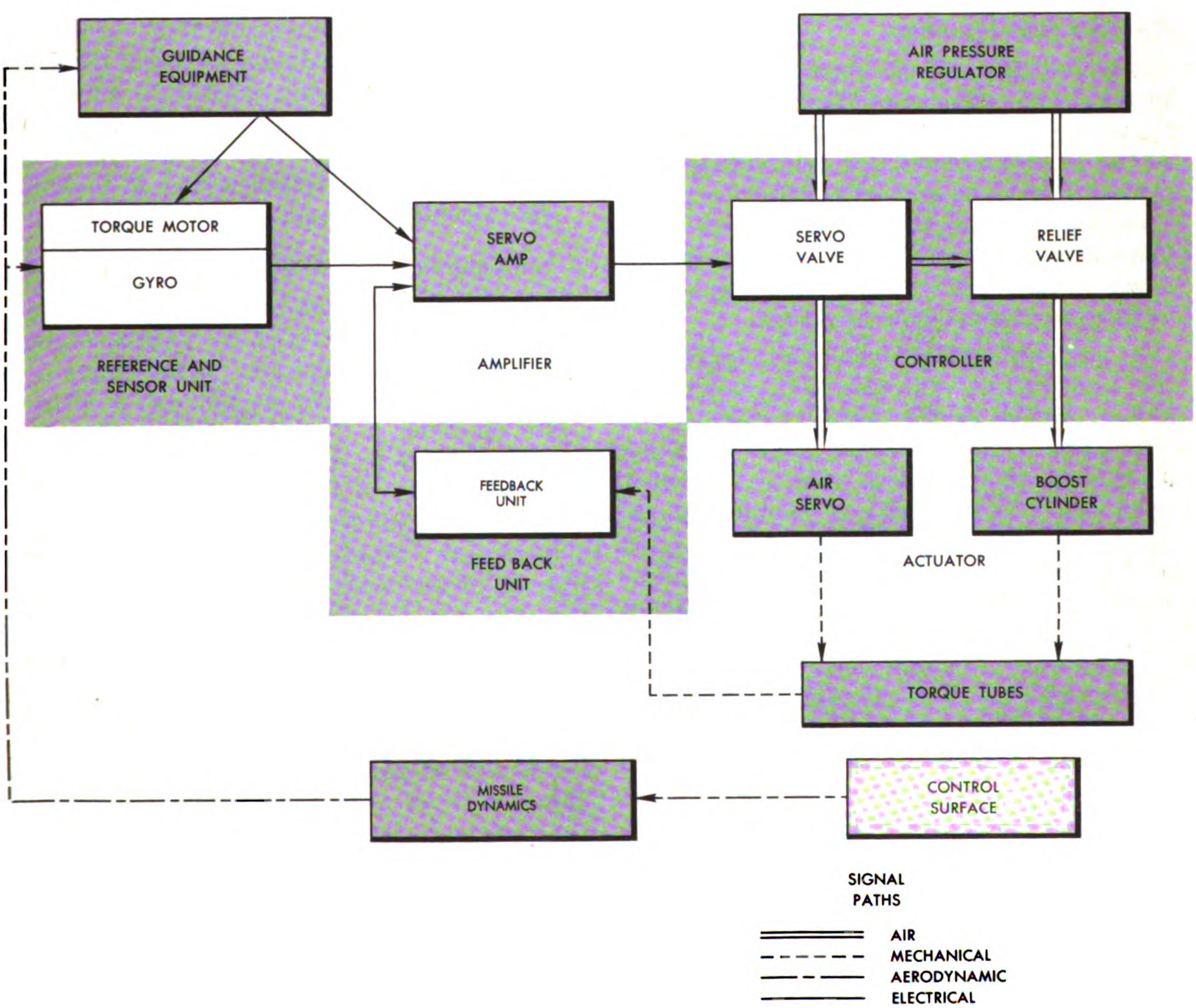
Remembering the disadvantage of a pneumatic system, you should realize the fact that both missiles are small and slow in speed is not just a coincidence. Air is compressible; consequently, time is required for air to

create a pressure in the piston sufficient to move control surfaces. Therefore, an air system responds slower than a hydraulic system. Although an air system is fast enough for slow missiles, it would not likely be used on high-speed missiles. Furthermore, in the section on aerodynamics, it was explained that more *force* is required in order to move control surfaces on large or high-speed missiles. The cushioning effect of air prevents it from being the best medium through which to exert the large forces required to move control surfaces on large, high-speed missiles.

This disadvantage is partially counteracted by the addition of a booster device as illustrated in the block diagram of a pneumatic-electric system below. When the servo valve moves a certain distance from midposition, air is allowed to escape through ports into a relief valve. The relief valve admits high-pressure air to the boost cylinder. The boost cylinder aids the actuator in moving the torque tubes and wind flaps. The boost cylinder is used to provide the additional force required to obtain large deflections of the wind flaps in either direction.

General Operation of Pneumatic-Electric Control System

Again, a gyro provides a dependable and accurate attitude reference for the missile. The gyro in a pneumatic-electric system is a viscous-damped integrating gyro.



Pneumatic-electric control system

The sensors are electric reluctance pickoffs which detect gyro displacement and produce a voltage which is proportional to the angle of missile-heading deviation. This output must be amplified in voltage and power so as to operate a solenoid and air servo valve. As in any control system, the servo amplifier performs this job.

At the air servo valve, the system changes from electrical to pneumatic. The air servo rotates the torque tubes. These torque tubes connect to the wing flaps and extend into the center section of the airframe. They impart servo motion to the wing flaps. The preceding block diagram applies to either pitch or azimuth control since the operation is the same.

Guidance Signal Insertion

The gyro error signal is not the only signal entering the servo amplifier. An electrical followup signal combines with it in the input to the servo amplifier. Also, a voltage from guidance could be applied at this point. The computer block of the general control block diagram is represented in this system by the mixer circuit, which combines these signals in the proper proportion and sense.

A guidance voltage could be applied to the control system through the gyro torquer-motor instead of the servo amplifier. The alternative inputs are shown in the preceding block diagram. A certain voltage dependent upon the guidance change required would operate the torquer motor to exert a force on the gyro. The torquer motor would precess the gyro at the desired turning rate. As the gyro was precessed and as the error pickoff detected the gyro displacement, movement of the missile would be produced. The gyro would maintain a constant angular displacement. The displacement would equal the amount the missile is off from its previous desired heading during a steady turn.

Due to target considerations, the guidance system in an interceptor-type missile must be much more complicated and pronounced than the preset guidance equipment in a JB-2. The guidance signals entering the control system would probably be produced by some type of radar and computing gear or by command guidance. The guidance signals override the gyro attitude error signals which exist at the same time.

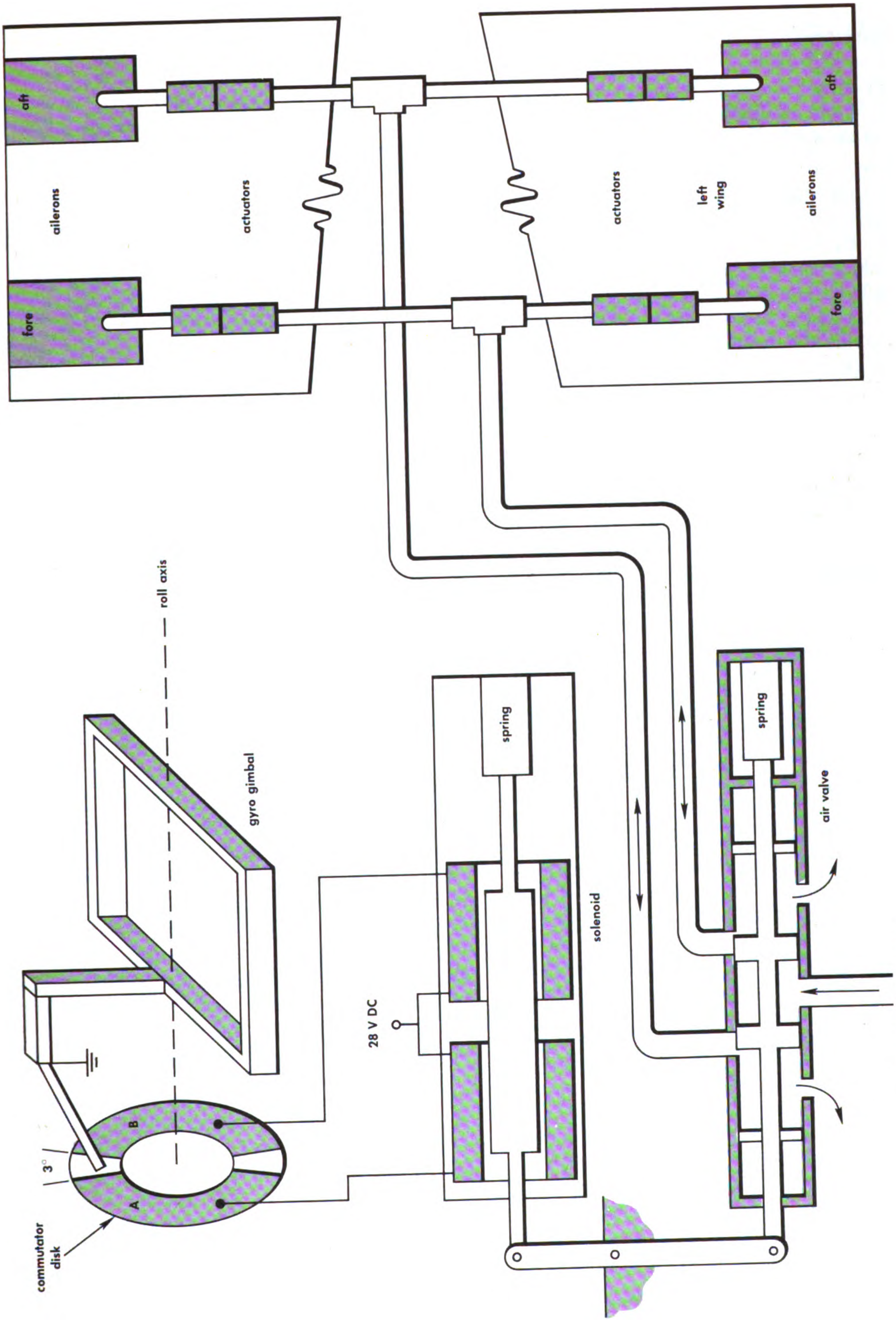
Followup

The followup voltage, derived from an autosyn mechanically coupled to the wing-flap torque tube, is proportional in amplitude to angular flap deflection. The voltage has a definite sense depending on the direction of flap deflection. Followup voltage is fed back with a sense to oppose the control input. This feedback cancels the controlling input from the wing flaps servo amplifier when the surface deflects a certain amount. The followup, then, produces a *deflection* of the wing flaps which is *proportional* to the input signal. The deflection for a given input signal must, of course, be a certain amount.

The outer or aerodynamic loop of the servo system indicates that wing flaps affect the position of the missile (missile dynamics), and this position change will in turn have an effect on the gyro reference and on the guidance signals entering the control system.

Roll Control

Roll control in the pneumatic-electric control system is achieved by a "bang-bang" (on-off) system, illustrated in the diagram at the right. As you can see, two sets of ailerons are mounted in the wings. Normally, they are retracted to the wings. If roll control is required, one set of ailerons completely extends out of and beyond the wing. The ailerons have a constant *angle of incidence*, and the angle of incidence of one aileron will be opposite to the aileron on the other wing, which extends at the same time. This produces the proper corrective rolling moment. The other set of ailerons is for roll in the opposite direction. The roll control system operates so that only one set of ailerons is extended at a time. Notice that *no* boost arrangement exists in the roll channel. One reason for this is that the force required to extend the ailerons is not as great as for the wing flaps. The angle of incidence of the ailerons has already been built into the wings, and the air servos need not turn the ailerons against strong opposing forces created by air currents.



A roll-control system

Air pressure to operate the control system (as well as to displace liquid fuel in the tanks) is supplied by compressed air bottles, which last the length of flight. Air pressure to the system is maintained constant by a regulator.

The control system for roll stability is simpler than the pitch and yaw channels. No signals from the guidance section enter the roll channel. Furthermore, the electric-type pickoff that is illustrated shows that the signal coming from it is either ON or OFF. The circuit to the roll solenoids is incomplete when the missile is level in roll because the slider of the roll pickoff is touching an insulated segment of the commutator. When the missile deviates in roll, the commutator *disc moves* with the missile, while the gyro gimbal remains stationary in space. The slide wire then touches either "A" or "B" segment. This completes the circuit to the proper solenoid. No amplifier is necessary since enough power can be passed through the pickoff to operate the solenoids. Not needing an amplifier is an advantage since it cuts down on weight, expense, and possible trouble.

The system cannot be called proportional control because the correcting action does not increase in angular deviation. No followup signal is necessary. There is no such thing in this system as the amount of aileron extension depending on the amount of error. A set of ailerons is either extended or retracted. The only feedback loop is completed through the movement of the missile. If the missile should roll more than 1½ degrees from level, a signal extends the proper set of ailerons. The signal remains on until the missile returns to 1½ degrees from level. This motion then opens the circuit at the pickoff, and the ailerons are retracted by springs.

Such a bang-bang system can only be used in applications where fine control of the missile is not necessary. The system must have a certain angle in which no corrective action is produced (dead zone). In the example we're discussing, the angle is 3°. This is a disadvantage since the missile can vary 3° without any control. If the dead zone did not exist, the missile would oscillate from side to side as the pickoff alternately contacted segments "A" and "B."

Channel Interconnection

In the control system discussed here, there is a separate control system for each of the three axes of control. These separate systems, sometimes called channels, operate independently. A roll error has no direct effect on the pitch or yaw controls. An azimuth error does not affect the ailerons. Therefore, this missile will not make a coordinated turn because only the rudder controls the turn.

One way to maintain roll control of a missile without ailerons is to use differential action on the elevators. Differential action simply means that when one elevator goes up the other goes down. This would roll a missile, and in such a case the pitch and roll channels would then be interconnected.

This combination roll-pitch channel is produced by feeding roll error signals into the pitch channel at some point. The roll signals must combine with the pitch signals in the proper sense. An extreme example of channel interconnection is illustrated on page 386 in the diagram of electrical interconnections.

While pneumatic-electric control systems are used in present day guided missiles, a still more commonly used type of control system is the hydraulic-electric system. Hydraulic-electric systems are discussed next.

Hydraulic-electric Control Systems

The combination hydraulic-electric (or electric-hydraulic) control system is used more than any other type of system in guided missiles. It is similar to the pneumatic-electric system. Like the pneumatic-electric system, everything in the hydraulic-electric system is electrical except the controller and actuator. The actuators, however, are moved by hydraulic fluid instead of compressed air. Since hydraulic fluid is practically incompressible and air is compressible, some of the disadvantages of a pneumatic system are eliminated in a hydraulic system.

A completely hydraulic system is not feasible; in fact, sensing units which operate hydraulically have not been created. In order to create an operative hydraulic pickoff, excessive force would have to be exerted against the gyro reference unit; this force would cause errors in the gyro. The power of a small signal could be amplified by using some type of hydraulic valve in which the low pressure would control other hydraulic fluid which is at a higher pressure.

The use of a computer which operates by hydraulic means would be highly limited by lack of flexibility in its performance and design. You can see that the probability of the development of an all-hydraulic system for missiles is unlikely.

In a hydraulic-electric system, hydraulic pressure is maintained by a continuously operating pump during flight. Since the fluid of the hydraulic-pressure system is used over and over, the operating time of the system is not limited. Thus the system can be used on long-range missiles.

Gyros act as an attitude reference for the missile. Electric pickoffs sense any missile deviations in relation to the gyro. The signals are computed and amplified by electrical means. The amplified electrical signals operate the controller which is some type of hydraulic transfer valve. The transfer valve regulates the amount and direction of fluid flow to the actuator.

ADVANTAGES AND DISADVANTAGES OF HYDRAULIC-ELECTRIC SYSTEMS

A disadvantage of a hydraulic-electric control system is the need for two sources of power: electric and hydraulic. This requires maintenance, assembly, checkout, and firing personnel who are trained in electrical and electronic systems. Also needed are personnel trained in hydraulic systems. Added to the personnel problem is the problem of supplying not only electrical parts but also hydraulic parts and test equipment to assembly, storage, and firing areas. Of course, this disadvantage exists in any combination control system. In the case of a hydraulic-electric system, the disadvantage is overshadowed by system advantages.

The most important advantages are the high speed of response and the large force available when using hydraulic actuators.

Also, a sensor and computer which operate electronically can be designed to produce an output of the correct form for every error which the missile may acquire in flight. A signal of the *correct form* is one which, at all times, is of the proper amplitude and sense to produce actuator movement which should result in the required correction. Couple this electrical system with a rapid actuator, which responds to signals the required amount regardless of impending forces on the control surfaces, and the result is an almost ideal system.

Again, one must look at the guided missile field from a broad standpoint. Keep in mind that a system which is satisfactory to one missile may not be as suitable for another type.

ROLL, PITCH, AND YAW CONTROL OF HYDRAULIC-ELECTRIC SYSTEMS

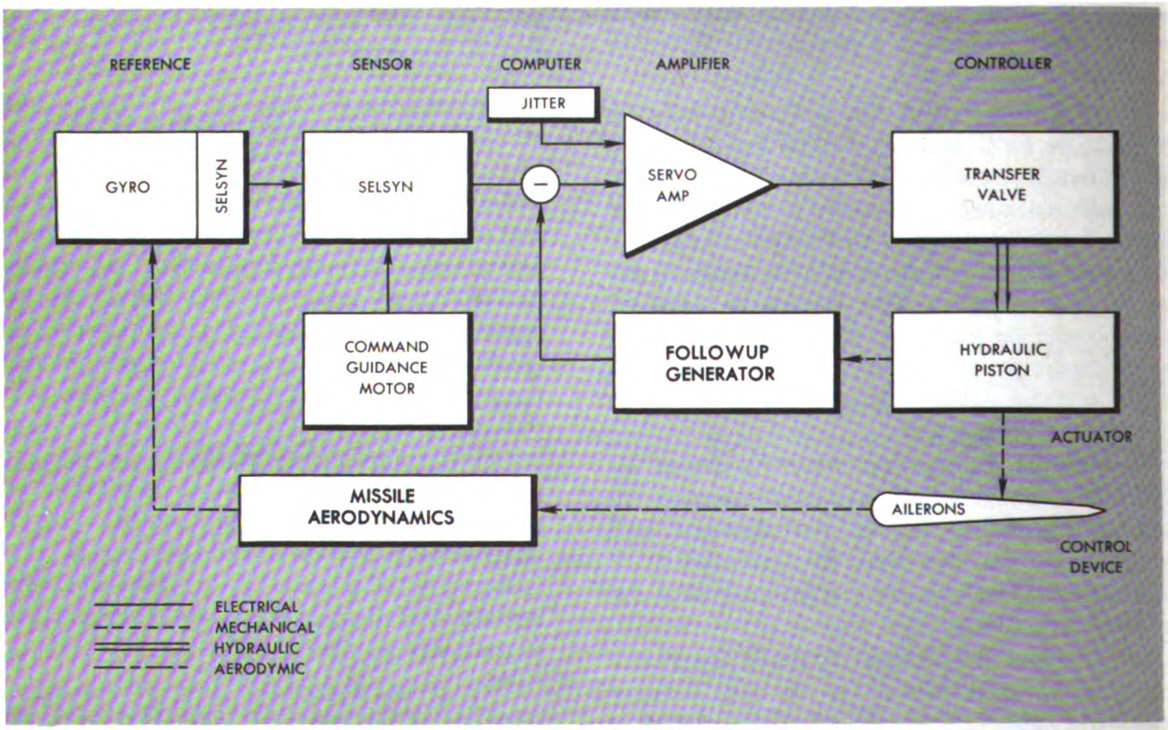
As in the case of pneumatic-electric systems, there are three distinct systems (channels) in hydraulic-electric systems for controlling missiles about their three axes. Let's consider these systems separately.

Roll Control

A hydraulic-electric system used to prevent a missile from rolling may operate as shown in the diagram at the bottom of this page. This system is simple, and it is used on a missile which has little tendency to roll in flight. In other words, the missile is aerodynamically stable in roll.

The system uses proportional control only; that is, it reacts to information which tells the amount of variation of the airframe from level flight. To do this, the signal is proportional to the deviation and is called a displacement signal. In this particular system, *rate control* was found to be unnecessary. Rate action is explained in the preceding section.

Notice that the components can be closely related to the basic control block diagram which is presented in Chapter 5. The gyro acts as a reference for the system. The selsyn is connected to the gyro and senses any roll of the missile. If any roll is detected, the selsyn produces an error signal. The correction signal to the servo amplifier is the difference be-



Hydraulic-electric system for roll control

tween the followup signal and the gyro signal. This voltage is amplified and increased in power so that it can operate the controller. The controller is a hydraulic transfer valve which regulates flow of fluid to the piston, actuating the control surface. The hydraulic system, which is needed to provide steady pressure to the transfer valve, is covered in Chapter 5, section 5.

Now is a good time to explain some commonly used symbols which you will see from time to time in the illustrations of this and following chapters. These symbols, representing certain operations in the system, are used by control personnel to explain system operation. The symbols save time and space.

The circle enclosing an "X" indicates a connection or mixing circuit for several signals. Signals are added if the circle contains a plus sign; that is, the output signal is the sum of the several inputs. If the circle contains a minus sign, the resulting signal is the difference of the input signals. A triangle amplifier symbol forms an arrowhead which always points in the direction of the output. A followup generator could be any pickoff which detects position such as a selsyn, potentiometer, or reluctance pickoff.

A block labeled *jitter* represents a generator or oscillator which provides a small AC signal of about 25 cycles. This jitter signal keeps the transfer valve and mechanical equipment vibrating. It overcomes the friction which results when these parts are not moving. Jitter appears in almost all missile control systems.

The control devices are generally shown as ailerons in this manual. This would not necessarily be the case; spoilers, elevons, or controlled jets might be used instead of ailerons.

Pitch Control

A pitch-control system which is a hydraulic-electric combination is shown on the following page. In this system a gyro and selsyn are again used as the reference and sensor units for changes in pitch attitude. Several amplifiers precede the servo amplifier as shown by the appropriate symbols in the block diagram. The purpose of this pre-amplifier is to strengthen a weak selsyn signal before it is fed to the demodulator or servo-amplifier input. Part of the proportional error is fed to the computer, and the rest is fed directly to the servo-amplifier. The servo-amplifier increases the power of the selsyn signal to the point at which the signal can operate the transfer valve solenoids.

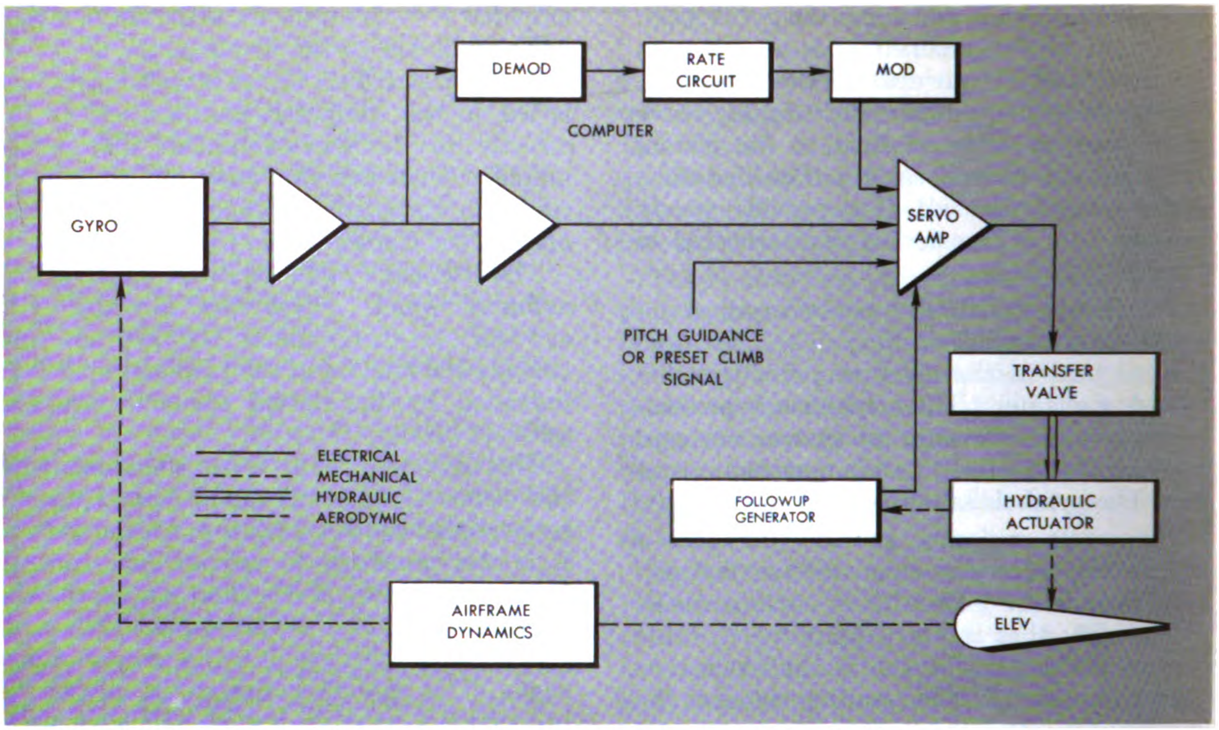
The transfer valve controls fluid to the actuator which moves the elevator. The elevator, in turn, affects the attitude and changes the relation of the gyro and missile as indicated by the dotted dynamic feedback path.

This dynamic loop, when combined with the electrical followup signal, completes the proportional control. In this particular system it is also necessary to provide rate control. Rather than use a separate rate gyro, the displacement signal is *operated on* by a computer which changes it to a *rate-of-change* signal. This rate-of-change signal affects the proportional signal in the same manner as in the pneumatic system. You can conclude, then, that the rate circuit is of the RC type since it is preceded by a demodulator and followed by a modulator.

An impedance mixer at the input of the servo amp combines the signals in the proper proportion and phase. The amplifier located prior to the servo/amplifier prevents any appreciable followup signal from "going backward" and entering the computer demodulator.

PITCH GUIDANCE INJECTION. One servo amplifier input in the following figure provides for the injection of another signal into the control system. If any signal is injected at all, it is either a pitch guidance signal or a preset signal which makes the missile climb just after boost phase.

A guidance signal would be used as a corrective measure in an interception-type missile where yaw and pitch location is critical throughout the flight. This guidance in two axes would apply in surface-to-air and air-to-surface missiles. It also applies in surface-to-surface missiles during the terminal flight phase. The signals could originate from one of many types of guidance systems such as beam rider, command or inertial, or from some type of homing system. The gyro and guid-



Pitch hydraulic-electric system

ance equipment output would be operating together in the control section. The *gyro* signal would guarantee *stability* while the *guidance* signal would provide corrections to the gyro *heading*.

A preset signal is normally used to climb a surface-to-surface missile which is launched close to horizontal. Without this signal, the pitch channel would direct the missile in level flight after the boost phase, and the missile would not reach the desired midcourse altitude. The signal is preset by adjusting a barometer on airspeed control, unbalancing a pot, or displacing a selsyn rotor a given amount. In any case, the signal produced is just enough to provide the correct angle of climb. The desired angle of climb depends on such factors as thrust, aerodynamic characteristics, and the type of guidance system.

In addition to providing this climbing signal, provisions must be made to reduce and eliminate it, so that the missile will fly with level attitude during the midcourse phase. This can be done from the missile by using a programmed method of signal reduction. A timer steps a contact along a tapped resistor at prescribed intervals, and a motor drives a potentiometer contact or selsyn rotor back to null point at a given rate. These actions produce a pitch climb signal which slowly decreases to zero and slowly levels the missile at the cruise altitude.

Yaw Control

The diagram on the following page illustrates a yaw control system which uses hydraulic-electric power. The system for yaw stabilization is similar to the other two systems except for the addition of integral control. A reluctance pickoff produces a signal proportional to the angular displacement of the missile with respect to a directional-gyro sensor. This signal has three paths to the servo/amplifier. The middle path provides the proportional signal, the outer path rate control, and the lower path integral control. These signals combine at the servo/amplifier to provide the desired correction signal no matter how far, fast, or long the the missile should deviate from the desired heading. Two probable places are shown in the figure for injection of heading signals from the guidance equipment.

INTEGRATOR ACTION

The purpose of an integrator in a hydraulic electric system is to detect an error of a certain sense that has existed for a comparatively long period of time. This is done by producing an output which is not only proportional to the amplitude of error, but also to the length of time the error has existed. The integrator actually accumulates the error signal. There are different types of integrators which can produce such a signal. This signal is then mixed with the other error signals.

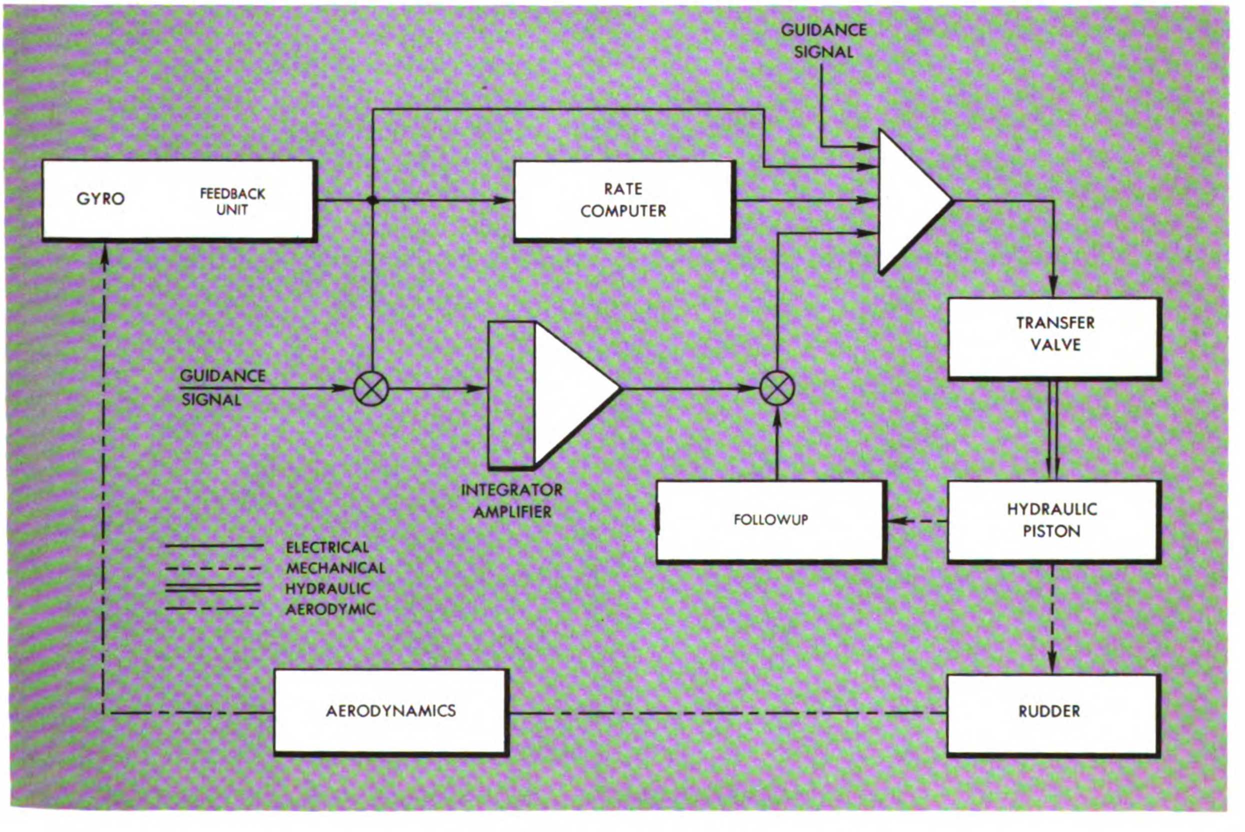
In the diagram below, the signal entering the yaw servo-amplifier is a result of displacement, rate of change, and accumulated error signals. The integrator is used in the system as follows: Assume that a missile was blown off course by a strong crosswind and that the control system produced a response which was insufficient for complete correction. The integrator responds to this small error and eventually produces a signal which is large enough to cause complete correction of the error. The integrator signal aids the proportional signal to overcome an error which exists for a period of time. It does not respond to rapid errors.

If you prefer, you may consider that, upon mixing, the integrator output changes the zero point of the more rapidly varying rate and proportional signals. The zero point changes because the integrator produces a comparatively steady signal, around which rate and proportional signals vary.

In certain circuits the integrator signal subtracts from the varying error signal. This would be used if a steady component of error signal is *not* desired.

CHANNEL INTERCONNECTION

As in other types of control systems, the hydraulic-electric system may have some type of channel interconnection. Connection between channels is necessary when control devices such as ruddervators, elevons, or controlled jets are used to direct the missile;



Yaw hydraulic-electric system

that is, interconnection is necessary when a control device must control the missile in more than one axis of movement. The systems covered so far have assumed the missile was being stabilized by control surfaces. This assumption is justified since the system principles remain much the same regardless of whether control surfaces, jet vanes, moveable jets, or fixed steering jets are used. Movable jet control would more likely use channel interconnection and is therefore used as an example here.

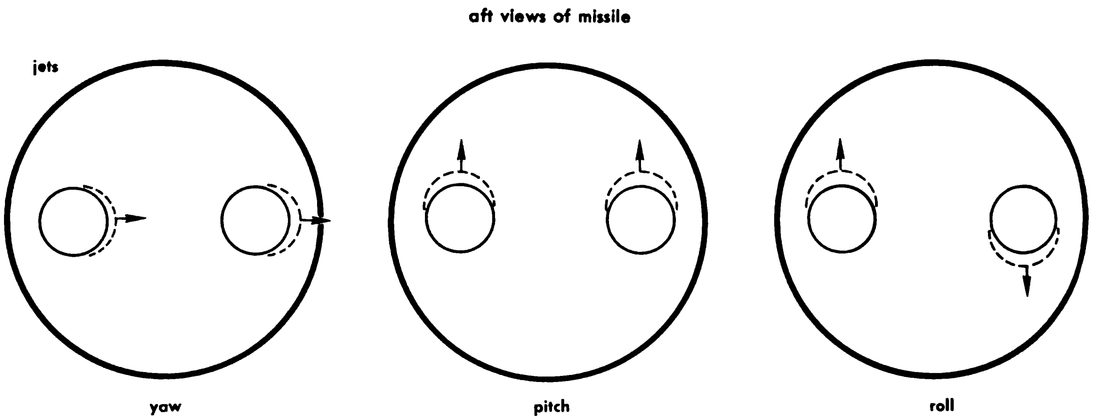
In order to obtain full control in all three axes of movement, at least two movable (gimballed) jets must be used. If only one jet were used, it would have to be mounted in the center of the aft section to produce straight flight, and it could provide no thrust to control roll of the missile. If two jets are mounted as shown below, it is possible to obtain full control. The jets must move in any direction, and each jet has to respond to signals from any of the three control channels

Let's consider a sample system which has four movable jets for stability instead of the minimum two. Each of these jets moves back and forth in only one plane. Two jets are used for pitch and the other two for yaw. All four jets operate for roll control. The roll channel must therefore be connected to both the pitch and yaw channels, a condition which introduces the problem of channel interconnection. If enough thrust for propelling the missile cannot be obtained from the use of these four rockets, another stationary jet can be mounted in the center.

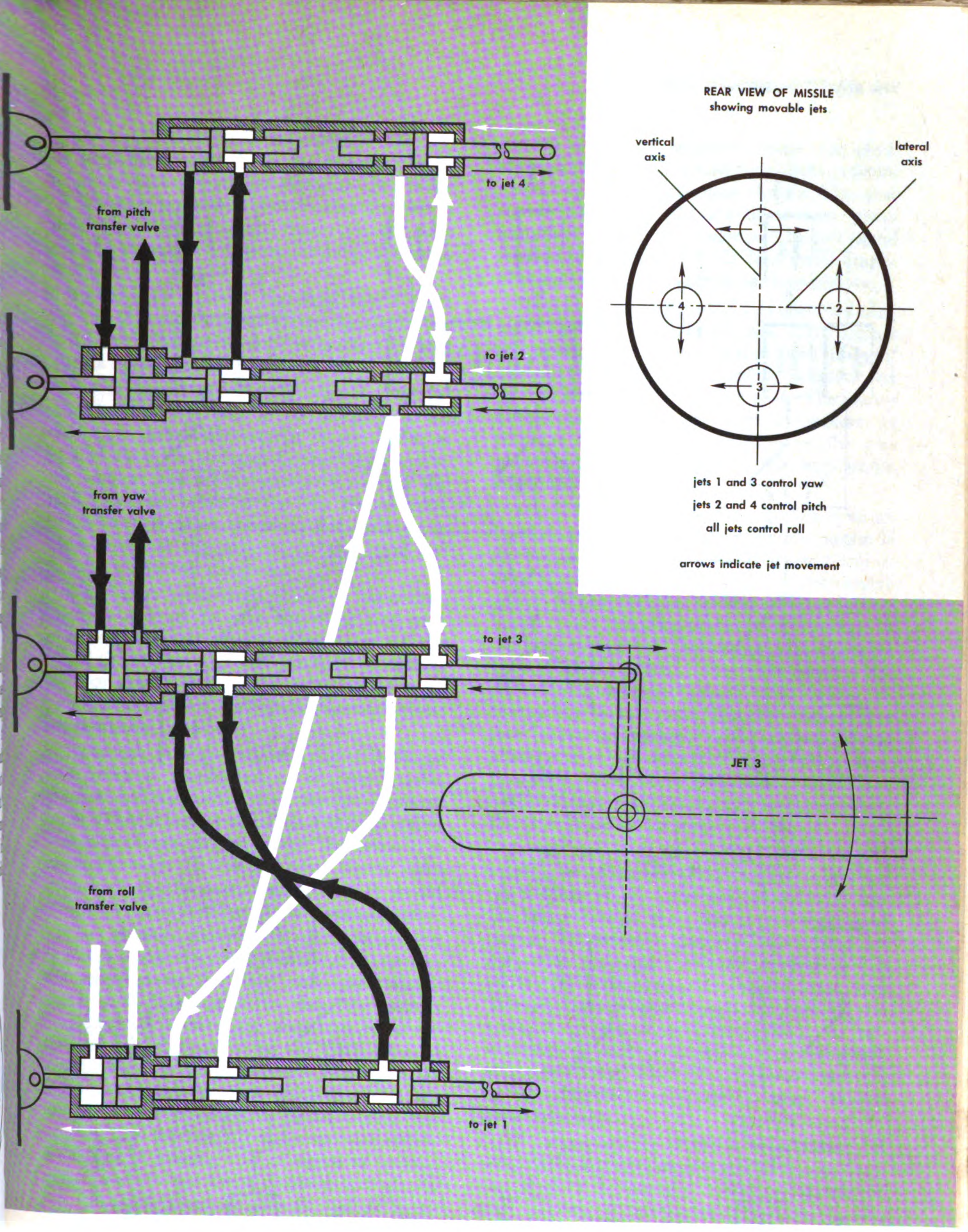
We will consider two methods of linking the roll channel to the pitch and yaw channels, one hydraulic and the other electrical. The first method, illustrated on the following page, uses three standard control systems up to the hydraulic actuators. Here the outputs of the *three* transfer valves are used to move *four* jets. The required differential movement of the jets is accomplished by hydraulic interconnection of the four actuators.

In the figure a certain fluid output from the transfer valve of each channel is assumed. Both sides of all the pistons contain fluid. However, for explanation purposes, the side which creates movement by means of higher pressure is shaded in color. Notice that the cylinders (piston housings) can also move. The arrows indicate either fluid flow, cylinder movement, or jet movement depending on arrow location.

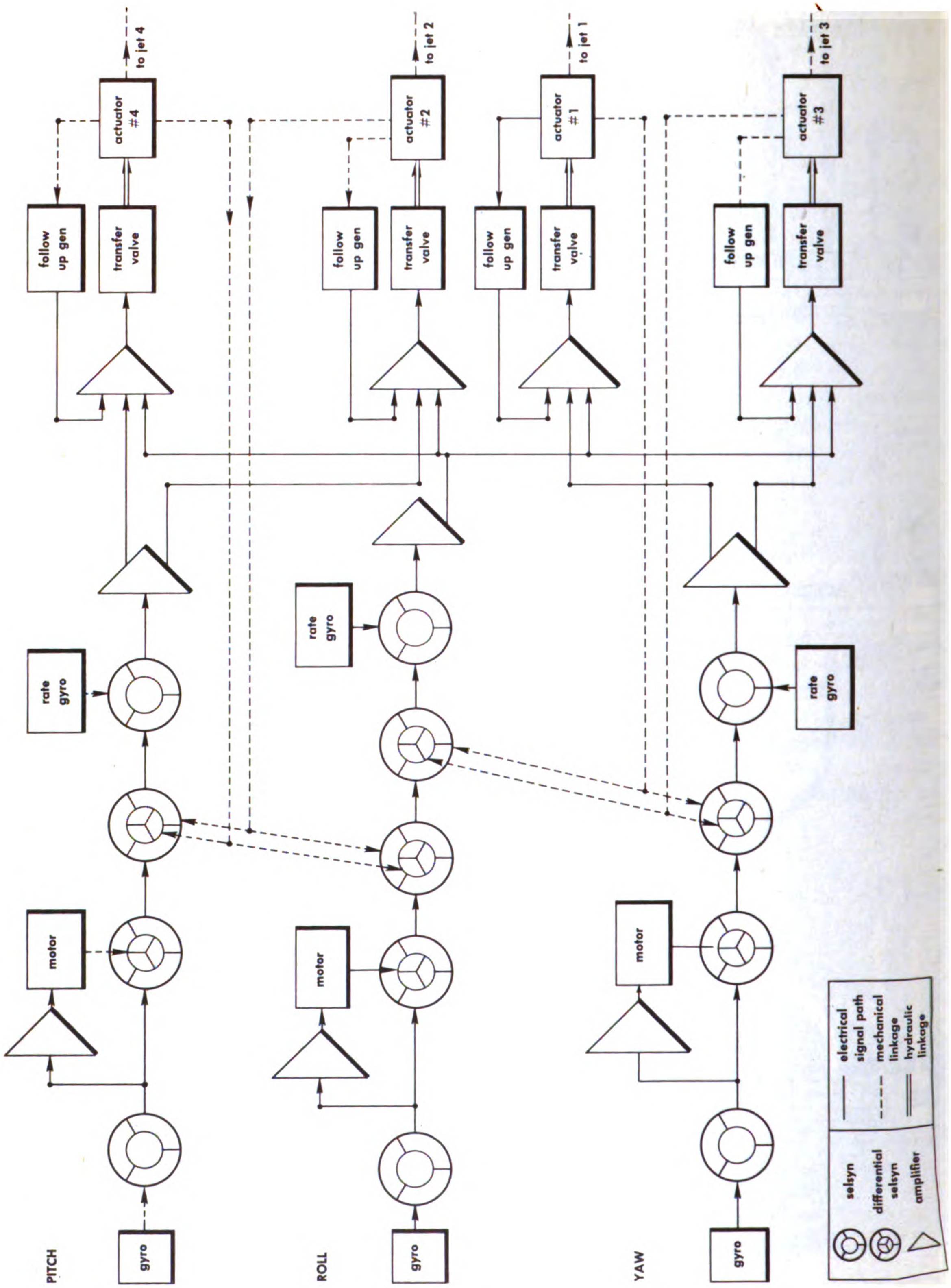
The method of obtaining hydraulic interconnection can be understood by assuming displacement of each transfer valve separately and then following the resulting flow of hydraulic fluid through the cylinders to determine jet movement. Assume that movement of the jet linkages to the right in the figure will produce clockwise movement of the exhaust nozzles as shown in the drawing representing the rear view of the missile.



Attitude control using two moveable jets



Hydraulic interconnection of channels



Electrical interconnection of channels

In the pitch channel, the assumed direction of fluid flow would force the exhaust nozzle of jet #4 clockwise and jet #2 counterclockwise. This action would tilt both rockets upward to produce a nose-up result. Since cylinder sizes are the same, both jets would move the same distance.

In the yaw channel, the fluid would force jet #1 clockwise and jet #3 counterclockwise. This would result in both jets aiding to produce a nose-right attitude.

A movement of the roll master cylinder would produce action in all jet pistons. In this case differential action is produced as one of the pitch jets goes down and the other up, and as one of the yaw jets goes to the right, and the other to the left. Hence there is no yaw or pitch reaction, yet each jet is aiding to roll the missile. In this case the missile would roll clockwise.

Note that while roll action is occurring, there is nothing to stop the pitch and yaw channels from producing corrections at the same time. The yaw and pitch action either aids or counteracts the roll action of each cylinder to produce the required instantaneous jet positions.

Note also that the followup signals indicate the relative displacement of the master channel pistons rather than the jet displacement. This is logical since the followup signal should indicate the result of the respective channels only. If the followup signal were connected to the actuator linkage, it would indicate the result of two channels in each case.

The second method of interconnecting channels is by electrical means. As shown in the diagram at the left, the interconnection occurs just prior to the four servo-amplifiers. The electrical distribution occurs by means of the output of the three channel amplifiers. The amplifiers are double-ended and produce two outputs which are 180° out of phase. The pitch signal again affects jets #4 and #2 by feeding into the respective servo-amplifiers. Assuming that a signal of certain phase produces clockwise movement of all the jets, then the signal to jets #4 and #2 must receive signals of opposite phase for pitch control. The double-ended amplifiers produce these required out-of-phase signals. The yaw signal also feeds into the other two servo-amplifiers in the same manner. The roll signal feeds into the input of all four servo-amplifiers, since it again operates all four pistons.

Rate and integral control has been included in the system, as you can see. Rate signals are obtained from rate gyros, and integral control from variable speed integrators fed by the displacement gyro signal. The many differential selsyns are used as mixers to combine position information from the displacement, rate integral, and followup selsyn generators.

Many followup paths exist in this system. In each case, actuator position information is fed to the input of the respective servo-amplifier. This feeding produces jet movement which is proportional to the servo-amplifier input. Also, actuator-position information is fed back to a certain two of the three channels (yaw, roll, or pitch). This is necessary because each actuator produces an effect on the missile in two axes. The roll channel has four followup signals, since each actuator affects the missile in roll. Thus, the combination of actuator followup signals to any channel produces a resultant signal which represents the true followup for that channel.

As pointed out at the beginning of the section, a hydraulic-electric control system (as is true of all combination systems) has the disadvantage of requiring two sources of power. The system places a heavy load on personnel training, because personnel are needed to work with two types of systems. Too, having to supply parts for both of the systems creates a problem. Even so, advantages of quick response and large available force have made the hydraulic-electric system the most frequently used missile control system.

The next section includes a discussion of electric control systems, which, although having their disadvantages, eliminate many of the problems of combination systems.

Electric and Auxiliary Control Systems

In this section we complete the coverage of guided missile control systems. Most of the chapter is concerned with electric control systems, but of no less importance are the auxiliary control systems covered in the last few pages. Coverage of auxiliary systems is included in this chapter along with electric systems only because of the few pages devoted to the auxiliary systems.

ELECTRIC CONTROL SYSTEMS

An electric control system consists of components all of which are powered by electricity. Thus, no pneumatic or hydraulic power-transfer system is necessary.

Except for the controller and actuator, the components used are similar to those used in the hydraulic-electric system. The heading reference is established by an electrically driven gyro and sensed by electric components such as a selsyn or reluctance pickoff. Any computing operation on the error signal is powered by electricity. A rate signal is obtained from the output of an electrically driven rate gyro or from an electronic rate circuit operating on the displacement gyro signal. Similarly, an integral signal is obtained by a motor-driven selsyn or by an electronic integrator circuit. Voltage amplification is obtained by an electronic amplifier. And power amplification is obtained by electronic means or by using a dynamotor-type amplifier such as a motor-generator set or an amplidyne.

A controller is a device that uses a signal to vary the electric power which moves the control surfaces. In the case of an all-electric system, the type of controller used depends on the type of actuator it must control. Further mention of controllers is made in later paragraphs.

Actuators of Electric Control Systems

Electric power can produce mechanical motion by means of magnetic force. This means that either a solenoid or a motor must be used as an electric actuator. A solenoid normally will not produce enough force to move an airfoil, so motors are used.

It is not practical to apply the torque of the motor directly to the control surface by using the motor shaft as the control-surface pivot. Such a motor would have to be gigantic in size to exert enough torque to move the airfoils sufficiently. A large motor cannot be used due to its excessive weight.

A small motor running at high speed has the same power potential as a larger motor which runs at some lower speed. Therefore, a small motor is connected to the control surface through a reduction gear train. The mechanical advantage caused by the gear train results in a large torque exerted on the control-surface pivot. The motor is either a *constant-speed* motor, operating through a clutch, or a *variable-speed* motor.

ADVANTAGES AND DISADVANTAGES OF MOTORS AS ACTUATORS. The required fast rotation

of a small motor introduces a major disadvantage to an electrical system. Suppose a sudden veer of the missile should require rapid action from a control surface. This means that the rotor of the small motor must run at a high speed to produce the rapid response. A large amount of inertia exists when trying to suddenly reach a high rotor and gear-train speed from a standstill. This inertia opposes change in velocity and causes an undesirable lag in the control-surface response. The lag may be so great that the system either operates with insufficient sensitivity or with a tendency to oscillate.

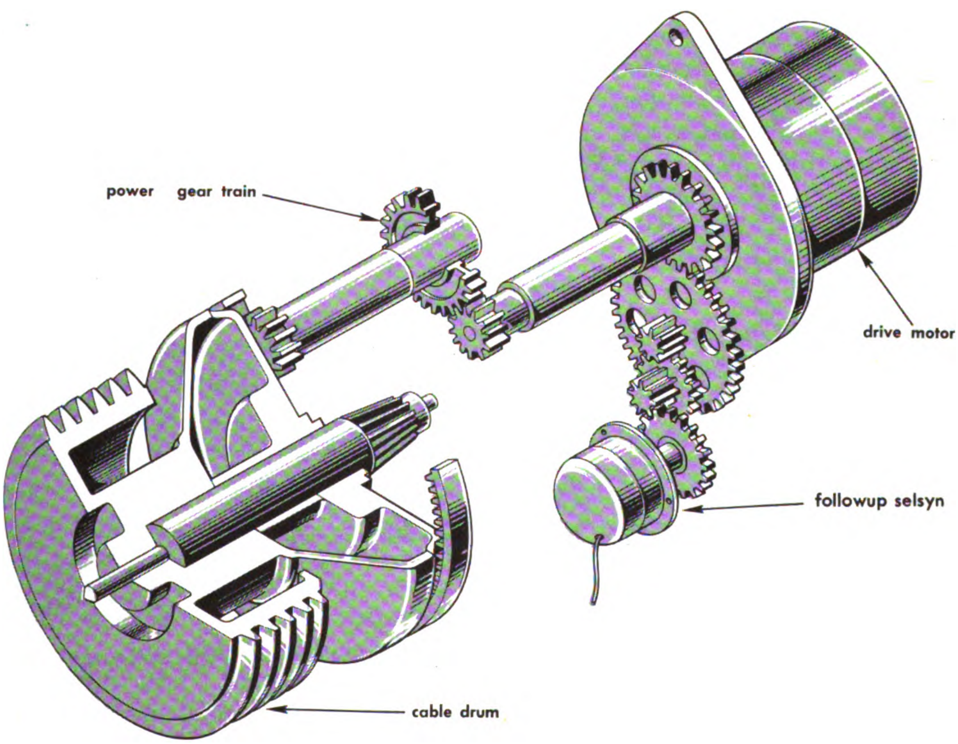
Even if a large motor could be used, the motor would still possess inertia due to its large mass. One means of partially removing this disadvantage is to control the output of a continuously operating motor by means of a clutch. A further improvement has been the development of small, high-torque motors.

The major advantage of an electric system is that the system uses only one medium (electricity) for operation. This reduces supply, assembly, personnel, and shipping problems as mentioned previously. Weight is reduced since only one power source is necessary.

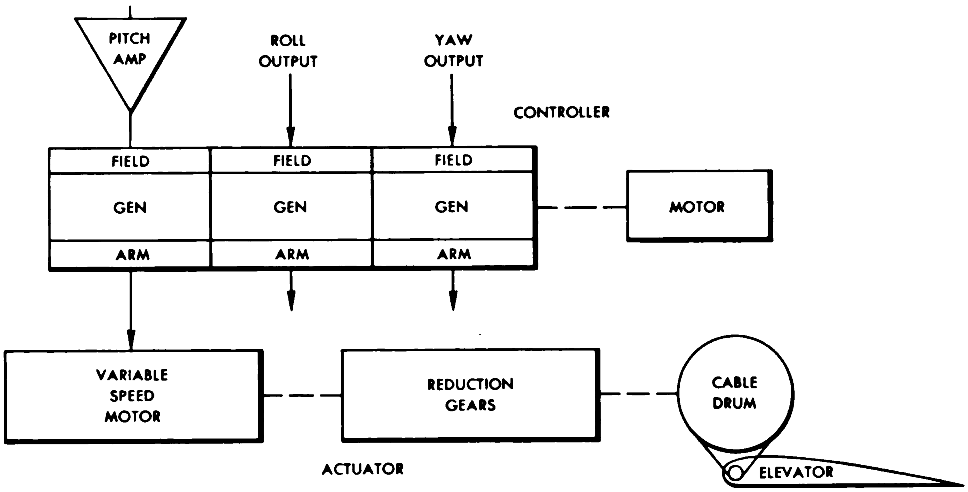
VARIABLE-SPEED ACTUATOR. The accompanying figure illustrates the use of a variable-speed motor to move a control surface. A signal is sent to a motor which rotates in either direction depending on the sense of the signal. Furthermore, the motor turns at a speed which is roughly proportional to the strength of the signal; the motor accelerates rapidly. Since the motor is coupled to the elevator through a reduction gear train, the elevator movement is proportional to the speed of the motor.

The signal to the motor must be of high power since it supplies the driving power of the motor. The controller must be capable of large output power. In some systems the variable-speed motor could be driven directly by the output of an electronic power amplifier.

The controller shown as part of the system on the following page converts the power of the controller drive motor to the three-channel variable-speed motors. Suppose the output of the pitch amplifier increases. The increased



Electric actuator (variable speed)



Part of pitch electrical system

output increases the magnetic field of the pitch generator. The voltage output of the generator increases and feeds power to the variable-speed motor which begins to turn. The transmission of this power puts an additional load on the shaft of the controller drive motor.

The controller drive motor must maintain reasonable constant speed regardless of load. Otherwise, if the pitch output decreased, the speed of the motor and, consequently, the output of the roll and yaw generators would decrease at the same time. This speed variation would cause undesirable cross-coupling between channels since the pitch signal would affect the other channels and vice versa.

CONSTANT-SPEED ACTUATOR. The effects of inertia when starting and stopping a variable speed motor can be eliminated by using a drive motor which runs continuously and maintains rather uniform speed. In this case the motor is connected to the control surface through a clutch which serves as the controller for the system. The clutch varies the power transmission from the motor to the control surface. The use of two clutches and a gear differential would allow control in both directions.

The figure at the right shows a system output using clutches. The figure also shows one of the clutches used in an autopilot. The friction clutch discs make contact by means of a solenoid from the channel power amplifier. The amplifier needs to supply power only to operate the solenoids.

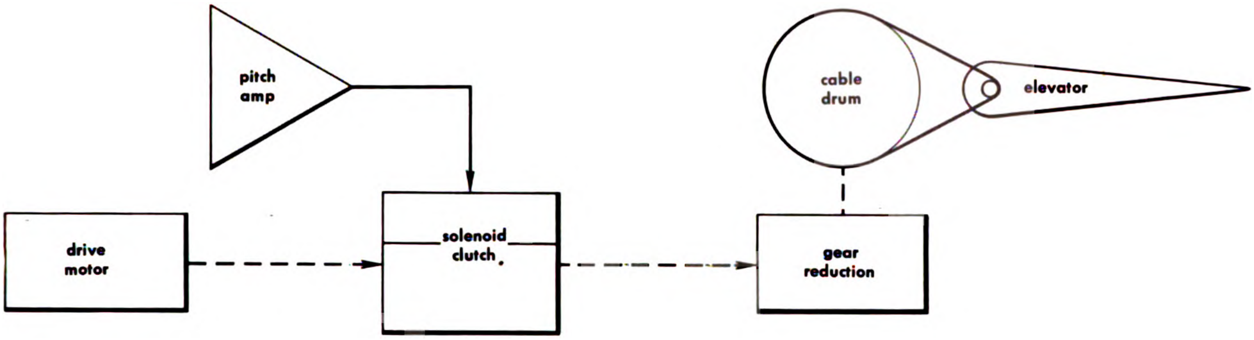
Use of the Dither Signal

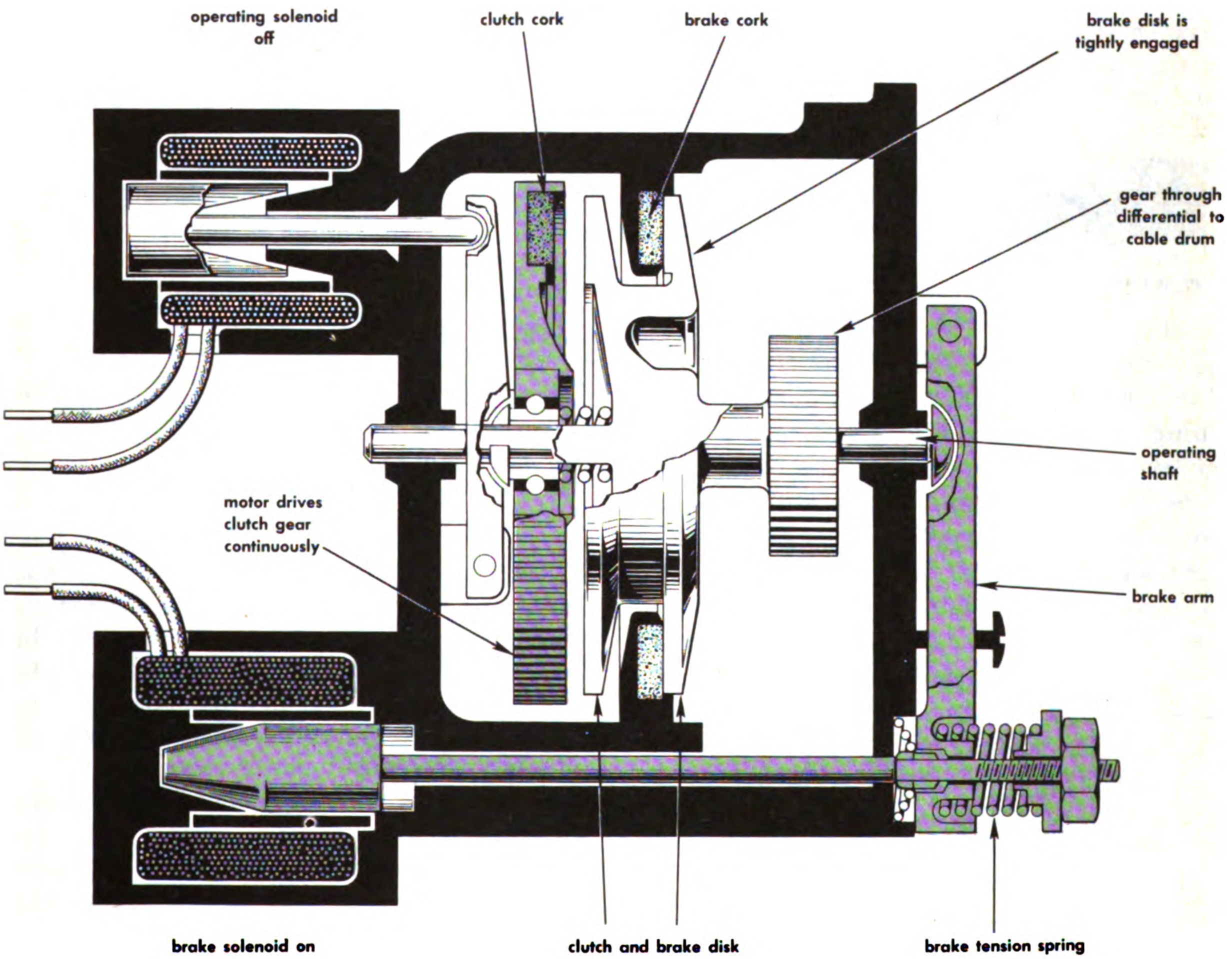
A *dither signal* creates small oscillations in a control system. This low frequency (about 25 cycles) is created by a signal from an electronic oscillator or generator, or it may be created by vibrating relay contacts. The signal is made to combine with the missile error signals.

The purpose of dither is to produce smoother action than that which the output of an on-off system would normally be. On-off control results when relays are used for controllers in an electric system. Normally such a system produces either full control surface deflection or none at all, since there is no way to distinguish large errors from small errors. The use of dither can produce action which depends on the amount of error signal as shown by the curves on page 392.

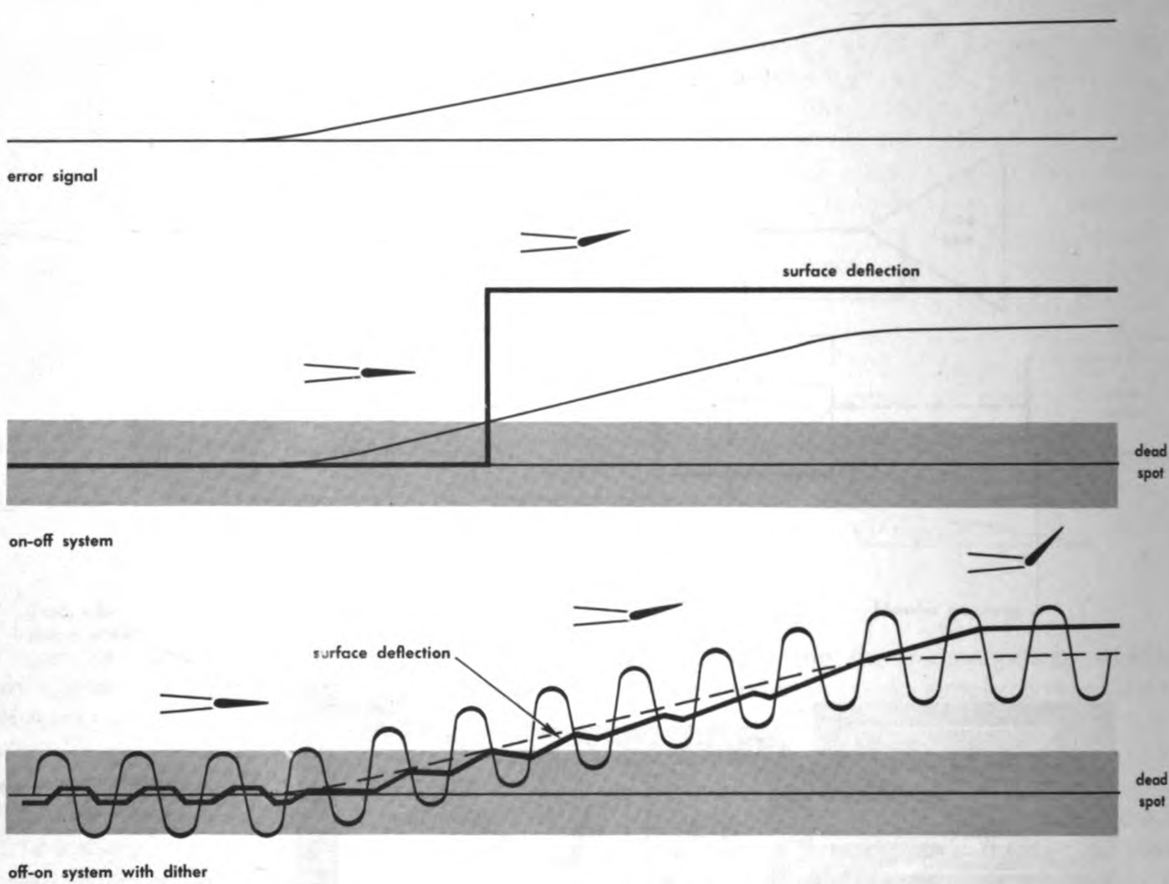
The top curve shows that the error signal starts from zero and increases. The middle curve shows that the on-off output signal moves the surface to full deflection when the signal becomes greater than the range of no action (dead spot). This dead spot is undesirable because small errors will not produce corrective action.

In the lower curve the rate of surface movement is dependent on the error signal strength.





Electrical system using clutches and motor-clutch control of autopilot



Smoothing action caused by dither

If no signal is present, a short output pulse during a half cycle of dither is canceled by a similar opposite pulse during the other half cycle, and a vibration of the airfoil about streamline position results. When even a small error signal appears, the time output in one direction increases and the surface moves further in one direction. As the error signal becomes larger, the time during each dither cycle in which motion is produced becomes greater. The surface then moves in larger jumps. This shows that the rate of surface movement is roughly dependent on the strength of the input signal. The on-off system, then approaches proportional control. a disadvantage of the system is that movement of the surfaces must be in steps rather than continuous.

This discussion of electric control systems rounds out the major types of control systems used for missile flight stabilization. The next few pages are concerned with control systems which do not have flight stabilization as a primary function.

AUXILIARY CONTROL SYSTEMS

Control jobs other than flight stabilization must be performed. They are performed by separate systems which may or may not be linked with stabilization control systems. The function of these auxiliary (supporting) systems is just as important as flight stabilization, although the putput power requirements are normally not as great. Accuracy requirements may be even greater.

Some guidance systems require that information be sent or received by the missile to a certain point via a narrow radio beam. The aiming of the missile beam relay antenna requires continual accuracy from an auxiliary control system. Any small antenna error caused by missile deviations could cause incomplete radio linkage.

Another antenna which requires an auxiliary control system is an unattended search radar antenna used in certain guidance systems. It also must be aimed accurately and must have a precise spin velocity. Also, the gyro slaving systems can be considered as auxiliary control systems.

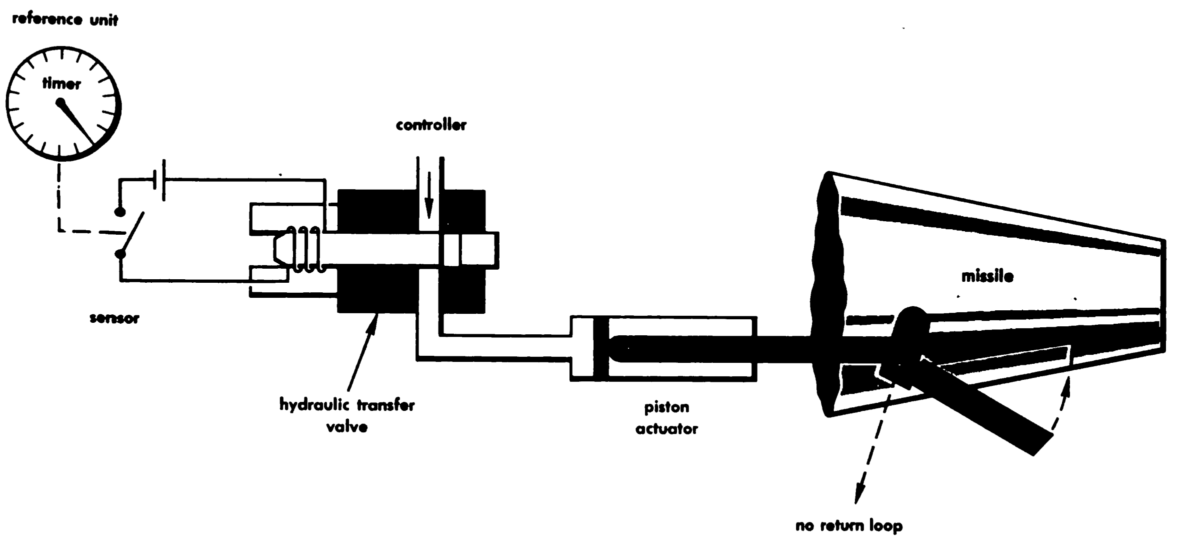
The operation of many of these auxiliary systems is similar to the operation of the stabilizing control systems which are given as previous examples in this chapter. An auxiliary system which, for example, maintains stability of a reference table is quite similar to a missile stabilization system. The system requires a reference and sensor unit to produce an error signal which indicates a comparison between existing conditions. This error signal is then changed in form and amplified to control an actuator, producing correction in the table position. This type of table is used to keep guidance sensing components such as telescopes or accelerometers in a certain steady angular position in space.

It is important for the system to be sensitive and accurate to keep the movement of the stable table to a minimum and to maintain it at a definite angular position. It cannot be allowed to tilt with the airframe. Any deviations may cause errors in the guidance sensing system or may even yield it inoperative.

Some of the auxiliary control systems do *not* operate on the same principle as stabilization systems. Stabilization systems operate on the servomechanism principle, which is a *closed-cycle system*. The closed-cycle system senses the results produced by the output and continually makes corrections. Many jobs to be done by the auxiliary control systems are simple enough for them to operate as an *open cycle* system. An open cycle system has no means of detecting the results and of making further necessary corrections.

These open-cycle systems are used to change some situation on the missile after it is launched. An example would be the closing of RATO doors after the boost rockets have been released. The illustration below shows a control system for doing this job. Notice that the major difference between this system and previous systems is the absence of relationship between the device being controlled and the input of the system. The action begins at the timer and ends at the RATO door, with no feedback path to the system to complete a loop. The absence of this complete cycle of action leads to the terms *open-loop* or *open-cycle* control system.

An open-cycle system could be used to reduce throttle setting from a terminal-dive setting. The system would set the throttle to



Open cycle-system: hydraulic RATO door closing

a slower position to prevent high-speed disintegration of the missile in its final dive. A limit switch connected to the throttle lever would cut off the throttle movement when it in in the new predetermined position.

The dive speed of the missile is not critical enough to require a servomechanism using a sensitive airspeed sensor to feed the throttle control system input for making further corrections. If such a system were used, it would then be a closed-cycle system with the missile speed providing a dynamic loop.

Two other uses of open-cycle systems are the erection of additional airfoils after the missile has cleared launcher obstructions and the deflection or protrusion of special flaps for the final dive. A terminal dive is straight down from high altitude and at full power till a maximum (terminal) speed is reached. The missile descends in a dive, but not a terminal one due to threat of structural failure.

The supporting systems which require high power output and fast response would normally be actuated by hydraulic means. As you know, this is true also in the case of the stabilization systems. Systems which do not have high performance demands would most likely be actuated by small electric motors.

These auxiliary systems are normally checked out by the same groups responsible for the stabilization control systems. One exception to this rule would be the checkout of the spin drive of the search radar antenna which is closely related to guidance equipment. Checkout procedures are, of course, similar to the stabilization system checkout procedures. Control checkout procedures are discussed in the next section.

Chapter 7 • Section 4

General Procedures for Checkout of Control Systems

Contained within any guided missile is a collection of electronic and mechanical parts which must be depended upon to replace a fighter pilot or perhaps a whole bombing crew. If a part should fail, there are no humans present to compensate for or to repair the failure, as is true in the case of guided missiles. Every part of the missile is important, or it wouldn't be there. Every part must function properly if the missile is to do its job. To reduce the possibility of failure during flight, the missile is checked as often as possible before flight.

The final prediction of reliability depends on preflight checkout. Therefore, the checks are designed to be as reliable and complete as possible. To be such, they must be made in the proper order and performed correctly. Checks which are performed incorrectly may result in inconclusive results and may even damage equipment. The manufacturer supplies checkout procedures to guide personnel on what to do and how to do it.

An important step in any checkout procedure is an accurate written record of the results. The technician, in cooperation with the mechanic, is responsible for gathering such information as step-by-step procedures are accomplished. Only the most important items are noted. Appropriate check lists are avail-

able. They vary in content according to the type and model of the particular missile. These records can be used as a basis for future improvement of the missile and as an aid for determining causes of flight malfunctions.

These checkout procedures may tell *when* certain checks or tests should be performed in order to assure that the missile is prepared for launch at the required time. Also, it is often necessary to coordinate with other checkout teams which must work on the missile prior to, during, or after a particular team is making checks.

For large missiles, the job is divided into major sections such as power system checkout, guidance system checkout, controls checkout, and airframe and propulsion checkout. Different teams are assigned for each division.

Military tactics and strategy command that the preparation for launching be short. Speed and certainty are the important objectives in preflighting a missile. The time lapse between the decision to fire and the actual launching must be an absolute minimum. For this reason, each preflight check is designed to be as short and simple as possible despite the conclusive nature of the tests. An assembly line procedure and a strict time schedule are normally used. A maximum amount of work, such as component and systems testing, may have already been done in storage and assembly areas and, of course, has been done at the place of manufacture. Obviously, an efficient and dependable checkout requires men who *know what they are doing*.

CLASSIFICATION OF TESTS

With the wide variety of missiles in existence, a large number of different tests and procedures exist. This makes it difficult to analyze checkout procedures from a general standpoint. However, one possible breakdown of all control system checks on different missiles is as follows:

1. Checks on the power supply system
2. Physical inspection
3. Electrical point-to-point checks
4. Zeroing and balancing adjustments
5. Servo static error simulation
6. Servo dynamic response checks

Although the six categories are listed in a logical order, it does not mean that they are necessarily performed in that order, or that the checks under any category are performed all at once. Some of the six categories may not even be included in the preflighting of a particular type missile.

The equipment which is included under the term *power supply system* includes those components which provide power to the control system and the actuators. Actuator power, remember, is hydraulic or pneumatic.

CHECKING THE POWER SUPPLY SYSTEM

The tests on the equipment which supplies power to the control system are normally carried out before the controls are checked. This is logical since the control system can not function properly if the correct electric, hydraulic, or pneumatic pressure is not maintained. For example, if voltage and frequency should vary, amplifier output would change. If hydraulic pressure is low, the speed of response of the system would be reduced.

In large missiles, the power supply systems are not checked out by control men but by other personnel with appropriate ratings. Briefings on electrical, hydraulic, and pneumatic power system checkouts follow:

Electric Actuation Checkout

The following general procedures must be accomplished:

1. Battery (if any) check

The battery must be checked for a full charge. Normally a new battery is used. The voltage regulation will be checked by measuring the terminal voltage under maximum probable load. The power wiring can be checked by measuring voltage at the load. Any high resistance wire or connections create an abnormal voltage drop when current is drawn.

2. Generator and voltage regulator test

The generator is normally driven by the propulsion unit, so any preflight tests on it will be timed to coincide with engine pre-

flight tests. The generator is normally used with a regulator which tends to maintain terminal voltage constant regardless of engine speed or changing load. The same checks for regulation and electrical connections would be made for the generator-regulator as for the battery.

3. *Controller check*

If any checks are made on an electric controller, they will be made for electrical balance and power gain, using specified input signals.

4. *Actuator check*

A main requirement is that the actuator move fast enough and with the proper sense with a given input signal. The actuator must be free from physical defects such as binding bearings and bent shafts.

Pneumatic Actuation Checkout

The following general procedures must be accomplished:

1. *Filling of air reservoirs*
2. *Pressure check*

Checks must be made at the reservoir and at prescribed points to guarantee that the pressure regulators are adjusted and functioning properly. This pressure may be critical in some parts of the system.

3. *Leakage test*

The presence of leaks can be determined by the ability of a system to maintain pressure over a period of time. Leaks can be located by applying rich soap suds around probable spots.

4. *Operating Test*

Tests must be made for electrical and pneumatic balance of the air relays (controller) and the air pistons. An air balance will guarantee that the piston will not creep off of mid-position when no signal is present. A mechanical balance will provide for equal movement to either side with opposite sensed signals of equal amplitude. Also, the friction of the pistons must be below a certain amount.

Hydraulic Actuation Checkout

The following general procedures must be accomplished:

1. *Filling and bleeding hydraulic system*

The written procedure for filling and bleeding the system should include miscellaneous precautions such as cleanliness, use of proper-type fluid, and information on the amount of torque to be used to tighten threaded fittings. Bleeding is the process of ridding the system of air bubbles, and it is normally accomplished by using a special test unit.

2. *Pressure test*

After the system is filled, it will be operated to see if pressure is maintained within prescribed limits.

3. *Leakage test*

Normally the system will be made to reach some pressure above operating pressure in order to guarantee that no leaks exist in any of the fittings or components. They can be examined visually after a specified period. The increase of pressure for this test creates an operating safety margin which aids in assuring dependable operation.

4. *Operating test*

During the above tests, the actuators will be operated to assure complete circulation of fluid. Tests might be required for mechanical and electrical balance of the controller. Component tests prior to final preflight could eliminate the necessity for this procedure. The actuators and transfer valve must be clean and possess a certain minimum amount of friction.

PHYSICAL INSPECTION

Some part of the control system checkout should include an examination of the physical condition of the equipment. Actually, an operator should be constantly watching for any indications which might show possible trouble. He may very easily see something which checks all right but may become defective during flight. Below is a list of things which can be watched for. Many of these items would actually be listed in the checkout procedures for a specific missile.

a. Frayed insulation or damaged wire
b. Poor contact of plugs in jacks.
c. Secure mounting of vacuum tubes.
d. Secure suspension of electronic chassis.
e. Cables retained in proper locations.

f. Breaks, cuts, or deterioration of rubber tubing.

g. Leaks in metal or rubber tubing and fittings.

h. Corrosion or dirt in equipment.

i. Worn or binding brushes in rotary equipment.

j. Dirty or rough commutator.

Some procedures include specialized checks for physical conditions such as:

a. The caged or uncaged condition of gyros.

b. The energized or deenergized condition of relays.

c. The operation of relays (visual and audible).

d. The position of switches.

ELECTRICAL POINT-TO-POINT CHECKS

Some time must be spent making electrical checks of the missile to determine if the steady state conditions are correct for launching. These checks are accomplished by making continuity, voltage, and resistance measurements between two points in the circuitry. The term point-to-point does not imply that the operator would probe the electrical equipment with test leads to make the checks. Better methods are used. Some of the conditions to be checked are supply voltages, plate voltages, and amplifier input and output voltages.

The condition of relays can be checked by making continuity checks through the contacts. In checking relay continuity, a contact is seldom checked individually. Instead, a series circuit is made up of normally closed contacts through several relays. This circuit completes a current path to a light on special test equipment. If the light is on, it indicates that all the relays involved are in the proper condition.

Another means of checking several relays in one operation is to connect the normally open contacts of several relays in parallel and let these contacts complete a path to a test lamp. If the lamp lights, it shows that one or more contacts are closed when they should not be. This alternative system has the drawback that a burned out indicator lamp would indicate all tests are successful.

Many of the circuits which are checked for resistance of continuity are miscellaneous independent circuits such as the arming circuit, destruct circuit, firing circuit, or parts of the command control or terminal dive circuits.

A seldom used method for making these checks is to probe the wiring with a multimeter. This method has serious disadvantages. Errors are easily made in locating the test points and serious defects can be introduced by probing the electronic equipment and rearranging parts. Errors can be made in reading meter results. All multimeters do not give the same results. Also, much of the electronic equipment is inaccessible, and checking by this means requires too much time.

These disadvantages can be partly eliminated by the use of a jack panel. This panel is wired to the missile electronic equipment and is mounted in an accessible location. By inserting test leads in different jacks, electrical contact is easily made to different parts of the circuit.

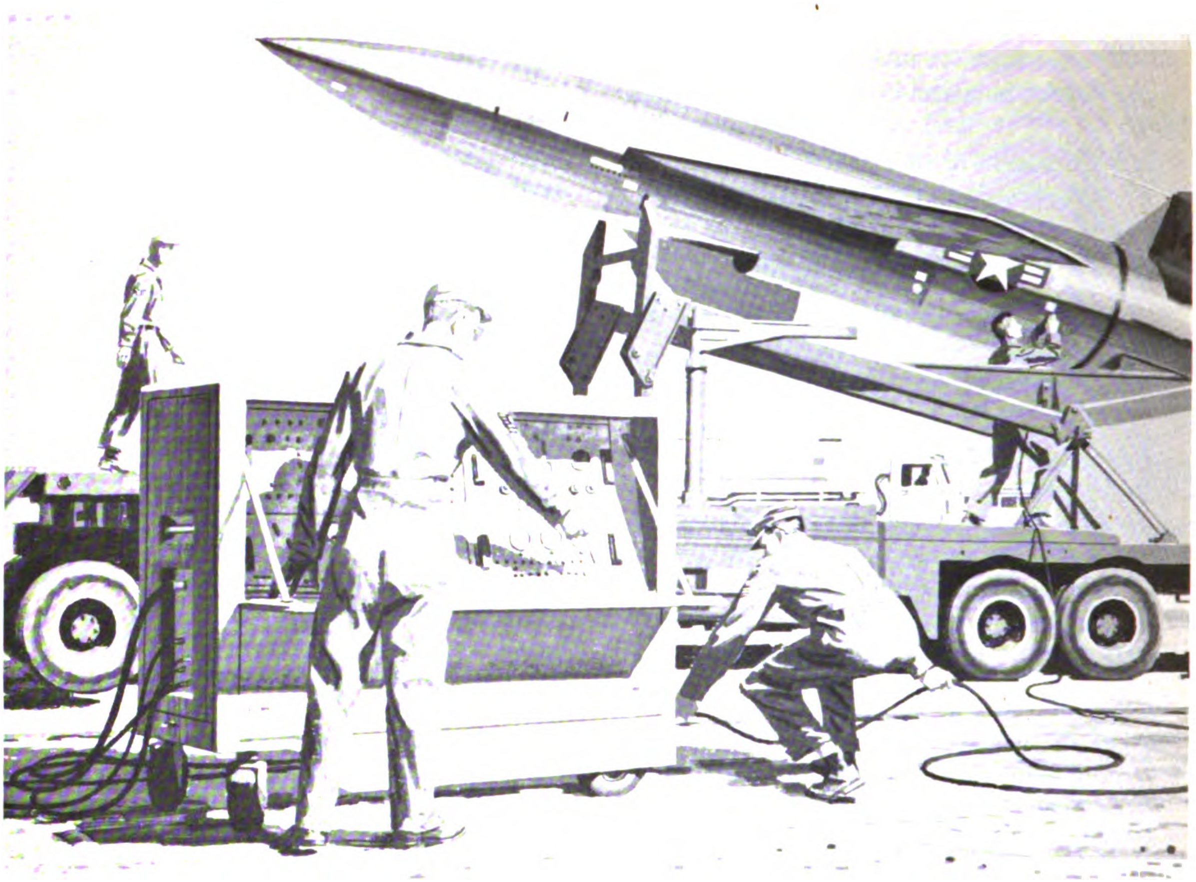
The tendency is to design test methods which are faster and more foolproof. This has led to the use of elaborate test consoles which make checks in a semiautomatic fashion.

Test Consoles

All of the necessary test equipment for testing the control system of particular type missile could be mounted in one rack called a test console. Another console could be designed to test some other section besides the control system.

These consoles are specially designed for certain type missiles, but form no part of the missile. Such a console is connected to the missile during testing by means of one or more sets of wire cabling. This cabling completes connections to all the circuits which are to be tested. The operator inserts signals to the system and measures signals from the system simply by moving switches and knobs on the console. Such a console may appear similar to the one shown on the following page.

The results are shown in some form on the test panel. They may be in the form of am-



Test-console checkout of missile

meter indication, a pattern on an oscilloscope, or the lighting of indicator lights.

Indicator lights are often used in making *go* or *no go* checks. Such a check indicates if signal strength is within the acceptable range by the lighting of a certain colored light. This eliminates a quantitative observation by the operator since a check is either *acceptable* or it is *not*. Such a system gives a definite answer in a minimum of time.

The major disadvantage of test consoles is that the console itself is complicated and is susceptible to malfunctions. Also, many of the adjustments in the console are critical, and maladjustment would lead to erroneous test results. Separate maintenance and check procedures for the console alone are established for this reason. To carry the job of testing to a still higher point of efficiency, test units are designed to test the missile test console.

ZEROING AND BALANCING ADJUSTMENTS

An important part of preparing a missile for launch is the zeroing and balancing of the control equipment. This has to be done before the checks which determine the operation of a complete system are made. Most of the zeroing and balancing is done on amplifiers or pickoffs. Many amplifiers which are used in servo systems have double ended outputs similar to the output of a push-pull audio amplifier. In some amplifiers this output must be balanced manually if the gain is to be the same for each section. The output must be balanced for DC plate current when this current is part of the output.

The pickoffs must be zeroed for the desired condition so they will faithfully indicate the amount of error. Those used in gyros are zeroed by the manufacturer, so the problem reduces to aligning the gyro case

properly in the missile. However, there are other pickoffs which must be zeroed, such as followup generators, airspeed transducers, and altitude detectors. They must be adjusted for the desired condition. All selsyns are adjusted to produce a null output when the rotor is in a certain position. The airspeed transducer is adjusted for a zero output at a certain speed.

Closely linked with the zeroing of pickoffs is the job of checking the linearity of the pickoffs. In most cases, the output of the pickoff should be proportional to the displacement. If a resistance bridge is used as a pickoff, a linear pot must be used. This pot has resistance which increases steadily with wiper movements. Linearity checks can be made far in advance of firing. There are other miscellaneous adjustments which have to be made on the equipment, such as adjusting timers or adjusting amplifier gain. The results of these adjustments can be observed on a test console.

STATIC SERVO RESPONSE CHECKS

The preflight duty of every pilot is to operate all of his controls to determine if they are operating correctly. He watches to make sure that the control surfaces are responding to his commands. If necessary, the aircraft could even be taken on a short trial flight prior to an important mission. When he gets into the air, his intelligence and ability will operate these controls to maintain stability of the airplane.

But a missile cannot go on a trial flight. Missiles are normally designed for *one flight — one way*. A missile does not possess the guiding influence of a human pilot but must rely on mechanical ingenuity and dependability for stability under all flight conditions. Since the missile cannot be flown for testing, the next best thing is simulated flight. There are three ways to do this.

If the missile is small, it may be mounted in a specially designed tilt stand for positioning at any attitude. By setting the missile at a certain angular position, the reference and sensor create an error signal, and the control system applies corrective action to the control surfaces. The surfaces move to a certain angle and remain in that position as long as the same error is maintained in the missile. At this point the followup signal is exactly cancelling the attitude error signal. With this attitude error, the angle of control surface movement must be within a certain range to guarantee proper response during flight. If the surface has deflected too far, the overcontrol produces instability in the missile due to oscillation. If the corrective action is too little, instability also is produced since the missile deviates too far before the correction is sufficient to return it to the desired attitude.

For example, assume that the pitch channel is being checked; the missile is tilted 4 degrees nose-up, producing an elevator movement of 6 degrees. If this is within the prescribed range, another trial with a nose-up attitude would probably be made but with a greater angular deviation. After this, the sense would be reversed by putting the missile in a certain nose-down attitude. The ratio of elevator deflection to missile deflection is often called *elevator gain*. After one channel has been checked, the other channels are checked in a similar manner.

A second method of simulated flight testing involves mounting gyros on a movable platform. This method would most likely be used if the missile is large. The reference table is removed and mounted on a tilt stand beside the missile for testing. All electrical connections from the platform to the missile would remain completed. Tilting the reference table, of course, simulates missile attitude errors.

A third method of simulating attitude errors is to connect artificial error voltages to the sensor. These signals are of prescribed amplitude and sense. The signals are equivalent to the output of the sensor when the missile is deviating at a certain angle. The reaction of the control surfaces is then measured. The control surface *gain* can be determined by the ratio of angular deviation over voltage input. A disadvantage of this method is that the sensor and reference units are not checked in the process.

The signal which is injected simulates a normal error signal from the sensor. It usually is AC and has a definite phase and frequency

in relation to the reference voltages of the control system. The signal is injected with several amplitudes, and the phase also is reversed.

When these checks are completed by any one of the three methods, it has been determined whether or not the control surfaces move the proper distances for any constant attitude error.

A further check must be made on most control systems to determine if they will react properly to guidance signals. Since the output of the guidance equipment is normally electrical, an electrical signal is inserted into the control system at the proper point to simulate a guidance error. The guidance error, of course, indicates that the position of the missile is either too high or too low, or too far to the right or left. The control surface movement must again be within a certain range for a given guidance error signal.

Test signals, which are varying also, are applied to control systems. The purpose of applying a varying signal is to determine reaction of the missile with a changing error signal. These are called *dynamic response* checks.

DYNAMIC SERVO RESPONSE CHECKS

The checks previously listed do not provide a complete check of a control system. Although constant errors have been simulated and the results have been measured, additional checks are needed to predict the stability of the missile in actual flight. Actually, during flight, any attitude error is not normally constant but is varying irregularly. The error is changing in amplitude at different rates depending on wind, thrust, and missile inertia.

Many types of systems are not ordinarily given dynamic checks in the field. On those which are not, either the checks are performed at the factory, or the static and component checks are relied on for prediction of stability. The component checks may include a dynamic response check on some part which has a large influence on the response of the whole system. Such components would include rate circuits, hydraulic transfer valves, and actuators.

The closer any checkout comes to simulating actual flying conditions, the more thorough and decisive are the test results. *Dynamic response* checks are used to test the missile under simulated operating conditions. The term dynamic means *moving* or *operating*, while *response* means *reaction* or *result*. The purpose of performing dynamic response checks on a missile control system is to:

a. Determine system reaction from certain simulated errors which vary at rates which may occur during flight.

b. Determine if this reaction is within prescribed limits to guarantee flight stability.

The fact that proper movement is attained with a constant error does not guarantee that proper movement will be attained with changing errors of different amplitudes. The question remains, "How fast do these control surfaces move?" Measurements must be taken to determine whether the control surface movement occurs too soon or too late, and whether this movement is too much or too little.

The difference between checking a control system with a constant error and one with a changing error is similar to the difference between making any static and dynamic test. *Static* means *not moving* or *not operating*. For example, measuring the voltage of a battery when it is not producing any current would be a static check. Measuring the terminal voltage of a loaded battery would be a dynamic check.

Constant Versus Changing Error Signal

The results of a changing error signal are different from the results of a constant error signal. Consider some instant in which a changing error signal is producing a control surface deflection. A steady error with the same amplitude as the changing error at that certain instant produces a different deflection. This difference is caused by factors which affect the time of response.

The control surface movement will *lag* behind the error signal that produces it due to the mechanical and hydraulic *reaction time*. Any action, regardless of what it is, requires time to take place. The action of a control system is no exception to this principle. The electrical, mechanical, hydraulic, and pneumatic reaction times which are present in

control systems cause control surface movement to lag behind the error signal producing it. Any slack in connections, or bending and expansion of the mechanical linkages in a system, also produces a delay. A hydraulic lag is produced by the expansion and compression of hydraulic lines and also by the time it takes fluid to flow through the transfer valves to the pistons. The lag caused by the compressibility of air has already been mentioned as a disadvantage of a pneumatic type system. In addition, lag is introduced in a system because of the use of integrators. In the electrical part of a system, any components which affect the phase of an AC signal voltage have some effect on the response of a signal. For example, all coupling and filter capacitors and inductances contribute to a time difference between the input and output.

These lags in a control system are objectionable since they do not allow the missile to correct rapidly. One method to reduce overall lag is by the use of rate circuits (lead circuits) or rate gyros. The output of this rate equipment is combined with the proportional signal. If the effect of the rate signal is great enough, the end result may even be a lead rather than a lag.

Meaning of Lead

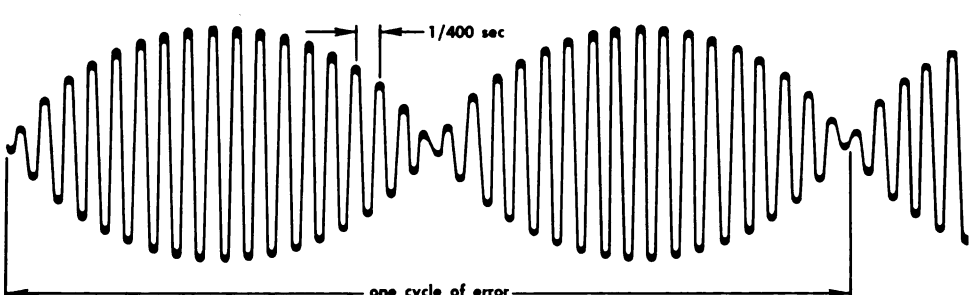
Does the term *lead* mean that a correction occurs before the error begins? That, of course, is impossible. Suppose the changing error signal for making dynamic checks is simply the sudden application or removal of a steady error signal. When the signal is suddenly changed, the system reacts an instant later to produce the new output condition. If a lead circuit existed in the system, the reaction would take place in a minimum of time. However, no actual lead could be measured.

If a simulated error signal, with the amplitude varying as a *sine wave*, is applied to a system, the control surface oscillates to each side of streamline (provided the error signal changes sense at each half cycle). By using this type of signal, a lead or lag could be measured. This type of changing signal is used most often for making dynamic tests. The frequency of this amplitude variation is low, somewhere between one to 15 cycles per second. In order to make the dynamic tests complete, several lead or lag determinations are made for every channel. Each check is made at a different frequency in order to simulate different rates of change of error. Because of this, the tests are sometimes called *frequency response* checks.

The deflection of the control surface also approaches a sine function. Whether this deflection leads or lags the signal depends upon the particular control system. The lead or lag is determined by comparing the *phase* of the control surface movement to the input signal after several cycles have passed. The amount of lead or lag must be within a certain range for each system. This range is determined by the manufacturer and is specified in checkout procedures.

Signal Simulation

Most control circuits are designed for AC error signals of a certain frequency, usually 400 cycles. The input signal must, of course, be modulated at this frequency. The most common method of simulating an AC signal, which increases and decreases with a sine wave function, is by rotating an autosyn at a rate equal to the frequency at which it is



Simulated error signal for dynamic response checks

desired to move the control surface. A constant AC voltage with a frequency equal to the frequency for which the control system is designed is applied to the stator. As the rotor passes through a null, the output decreases to practically zero and changes phase. It then increases with opposite phase. The average amplitude then approaches a sine wave function (see preceding page). The speed of rotation is adjusted to provide the modulating frequency.

Another method of simulating a changing signal is comparable to that used in simulating static errors. That is, the gyros, or the entire missile, are rotated back and forth about a certain axis at a certain rate by means of a support table. Such movement has to be provided mechanically if accuracy is expected.

Analysis of Results

After the changing errors are simulated, the final job is to measure the results and determine if they are within the required limits. The control surface deflection is best determined by connecting a pickoff to the actuator linkage.

This pickoff voltage is compared in *phase* and *amplitude* to the input signal to determine the response of the system. Voltage comparison can be done by four general methods:

1. Visual inspection on oscilloscope screen.
2. Reading of meters or oscilloscope connected to special phase and amplitude comparator circuits.
3. Inspection of paper tape record produced by pen-and-ink-recording oscillograph.
4. Inspection of record produced by photo-oscillograph.

One of the fastest methods is the use of oscilloscopes. Before the phase and amplitudes of the input and output voltage can be compared by this means, the 400-cycle modulated frequency must be filtered. Only the amplitude (or envelope) of this frequency is compared in phase. The phase and amplitude readings for different modulating frequencies are then compared to a table of limits. It is possible for this comparison to be done on the previously mentioned test console.

CONTROL SYSTEMS IN MISSILES

You have at this point completed this manual's coverage of missile control systems. First, you studied the basic components of control systems. Then this chapter introduced you to the systems as a whole. Now, having been briefed on control checkout procedures, you have a general understanding of what takes place in readying missile control systems for flight.

Guided missile control systems are all similar in operation regardless of the type of missile. Their purpose is to replace a pilot in coping with winds, equipment unbalance, and airframe aerodynamics so as to keep the missile stable in flight. The systems must also be capable of accepting location errors from guidance equipment and of making corresponding heading changes. The systems must respond sufficiently and rapidly enough to guarantee smooth, straight flight. For this reason, combination hydraulic-electric systems are most common.

While guidance systems differ greatly with types of missiles, control systems are similar, since standard servomechanism principles always apply. This fact makes the limited coverage of control systems and control checkout procedures presented in this manual more useful and more comprehensive than you might at first expect.

Chapter 8

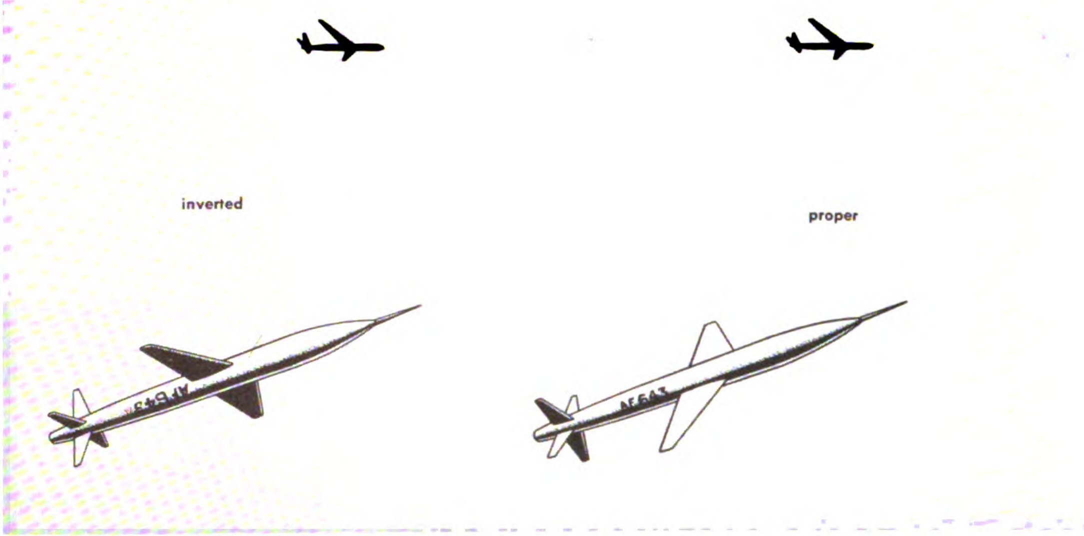
Trajectory Considerations of Guided Missiles

In this chapter, the path of a guided missile is considered to be a trajectory because guided missiles are "on their own" in somewhat the same manner as a bullet after it has left the gun barrel. Missile curve through space on a track which can be analyzed by advanced geometric methods. In the early stages of developing a missile, the possible trajectories that can be followed are rigorously calculated in order that the airframe and control characteristics will be tailored to the optimum performance of the missile under the expected conditions of its operation. Thus, it is not idle curiosity that prompts a study of missile trajectory but a necessity to the understanding of guidance requirements. For your use it is not necessary to obtain more than an *understanding of terms and basic principles;* you need not resort to analytic geometry.

REFERENCE CONCEPTS INVOLVED IN MISSILE SYSTEMS

A basic principle involved in all missile systems is the concept of the reference axes. In order to detect direction and magnitude of an error there must be a standard beginning point for up-down and right-left measurements so that the control system can be commanded to give the proper corrections. This standard becomes quite complicated as the complexity of the guidance system increases, especially for the complex system of long-range missiles.

A reference system can be based on an actual arrangement of some physical components, or it can be quite arbitrary. For arrangement of physical components, the homing-type missile could be cited as an example. The reference for such a system is the



Identical missiles in different attitudes making proper correction for targets which are above and to the side of missile

physical and corresponding electrical orientation of the antenna. In the homing missile, with a fixed antenna that looks directly forward, the arrangement is easily analyzed. Note in the sketches above, which show the effect of attitude on missile reference, that what is *up* to the antenna is *up* to the missile. Whether the attitude is inverted or proper, the command resulting from the error signal detected at the antenna serves to turn the missile toward the target.

In the initial planning of a flight, it is necessary to be exact about the primary reference of the trajectory coordinates. The selection of these coordinates in an inertia system is a rather arbitrary decision because the earth is pursuing its orbit about the sun and rotating about its own axis. The requirement, then, that the reference be stable or fixed in respect to the guidance system is not directly available. Since a gyroscope wheel maintains rigidity in space or, more scientifically, in inertial space, this inertial space becomes the reference. As far as inertial space is concerned, the reference is a particular direction, usually with the axis of the gyro parallel or perpendicular to the force of gravity at the time and place of launching of the missile containing the gyro.

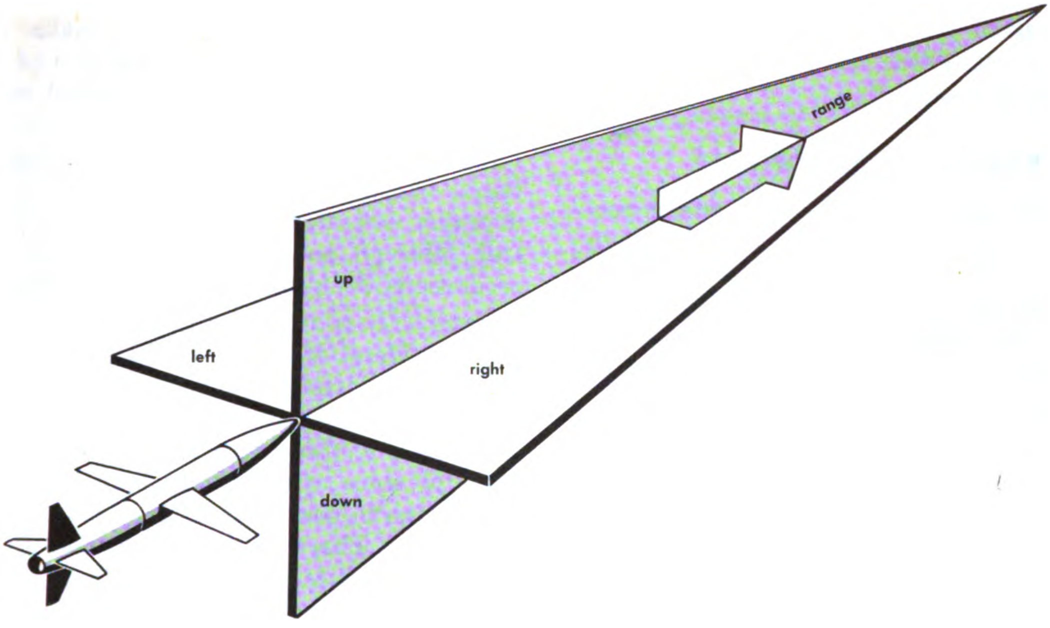
In this chapter's coverage of coordinate systems, you will be familiarized with these particular systems. You will be introduced to the variations and combinations of coordinate systems required by missiles for applications of the systems.

GUIDANCE COORDINATES

The simplest concept of guidance coordinates is the set already mentioned in which a seeking head is fixed to a missile so that its seeing axis is only to the front along the projected line of missile travel. The seeker head is orientated so that the signals going from its detector to the control system institute control movements that place the missile back on a path directed toward the target. When the missile rolls or changes attitude, the guidance form changes its attitude accordingly. Therefore, the guidance reference is fixed to the missile in the fashion shown to the right above. The guidance reference moves with the missile so that there is no confusion as to movement required to correct any error signal.

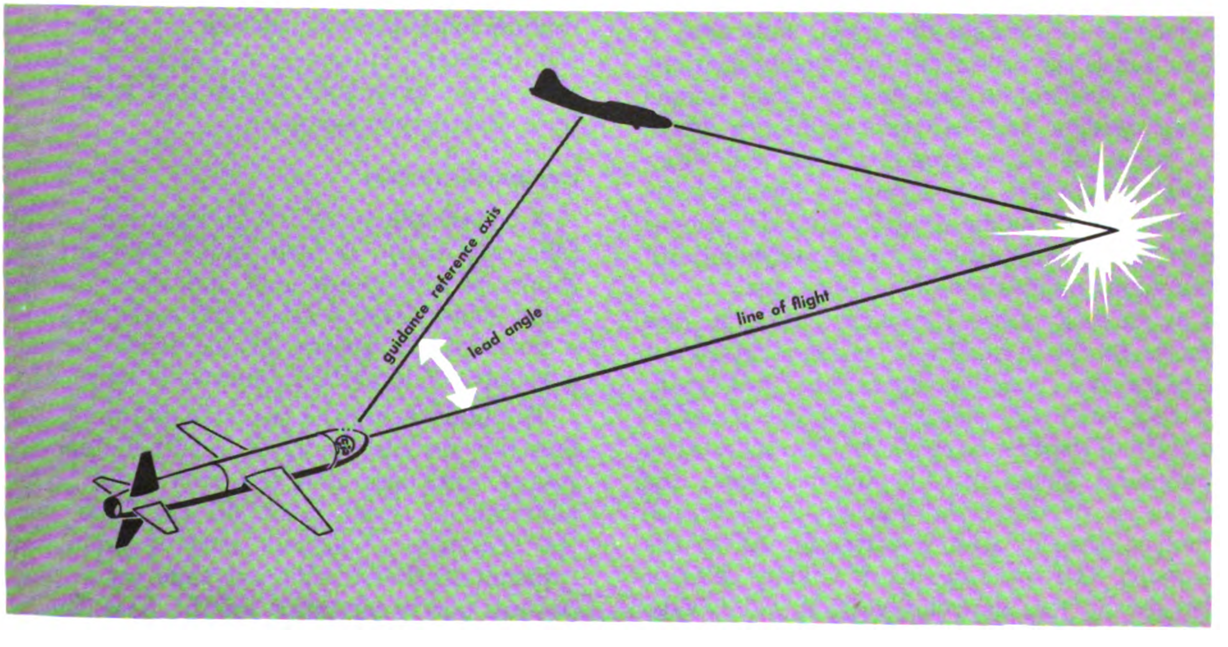
This could be termed a two-dimensional reference system since only up-down and right-left signals are detected. The range dimension may or may not be a function of the guidance equipment. In some cases range is an ordnance function only necessary for warhead detonation; it has no direct use in guidance in such cases.

In later developments the seeker head has been mounted onto a missile by a universal joint (gimballed) arrangement. With such an arrangement the seeker can look at a moving target while the missile course is projected at a lead angle or collision course, as illustrated on the right. The course is designed to make contact with the target with a minimum of maneuvering required by the missile. The



Simple guidance coordinates

missile "looks over its shoulder," and the error signals thus received are properly interpreted to the control system so that a correct collision course can be maintained. According to geometry, two intersecting lines determine a plane, which is a two dimensional surface. Thus, the problem could be considered two dimensional if the lines were ideally straight. The problem is made complicated, however, by the lead angle involved, requiring manipulation of the signal according to trigonometric functions. A common way to obtain trigonometric operations on an electrical signal is through the use of selsyns. The reference is still fixed to the missile but is transformed or interpreted from the antenna signal. The



Variable guidance reference

antenna signal is in the guidance reference to the control-system signal which actuates the missile properly.

TRANSFORMATION OF AXES

Just about all missile guidance systems perform transformation of a reference system to the axes of the control system. In some systems there are several of these transformations. In the system just outlined, the reference axes are rotated through an angle to make them correspond to the control reference. Rotation of axes is a means of transformation.

It is important to remember that when a transformation is performed on a system of axes or a coordinate system, the effect is on all three axes. You cannot tilt a side of a box without tilting the rest of the box (unless you destroy the box). The same is true of a three-dimensional coordinate system.

If there are two boxes, you place one box next to the box which is the standard in order to compare for size. To do this, you have moved the box. Moving a set of axes for comparison with a standard is called translation of the axes. Thus, transformation of a reference system can involve both rotation and translation. That a missile performs such an operation is not often obvious from reference to schematic diagrams, but such a function is the reason for many complications in the electronics equipment. The missile must receive and interpret guidance errors in terms of its reference frame, and compute the control movements in terms of the missile aerodynamic axes to give the proper corrections. The interpretation and computation steps require complex equipment.

The reason some guidance systems are chosen for use in missiles is because they do not require the complex operations just mentioned; not every missile performs a transformation of its error signal to develop the control command. Some missile systems detect the direction of the error and apply it to the control directly after a correction as to the amount. The computation of amount of error for use by the control system is not always a direct result of the concept of reference coordinates, but such a conception does enter in. Almost all missile systems make some attempt to distinguish the magnitude of error so a proportionate amount of control can be applied, thus the term *proportional control*. Even though the attempt to obtain proportional control is still crude, a time lag or damping circuit is a move toward such control. Without such a circuit, uncontrollable oscillations could result from large errors.

Externally controlled missiles require a further consideration. Such missiles have an ideal trajectory charted for them in space and are then commanded to stay on it. Either the flight path or the command is actually instituted by some means external to the missile. Remote radio control is an example of external control. The tracking information for the remote control is by visual, optical, or electronic means. Both the beam rider type of missile and the long range electronic navigation system, like the Loran system, utilize charted or projected paths of radio energy to guide them to their target. The desired course then is the prime axis of the reference system, and the missile is controlled so that it does not deviate from this.

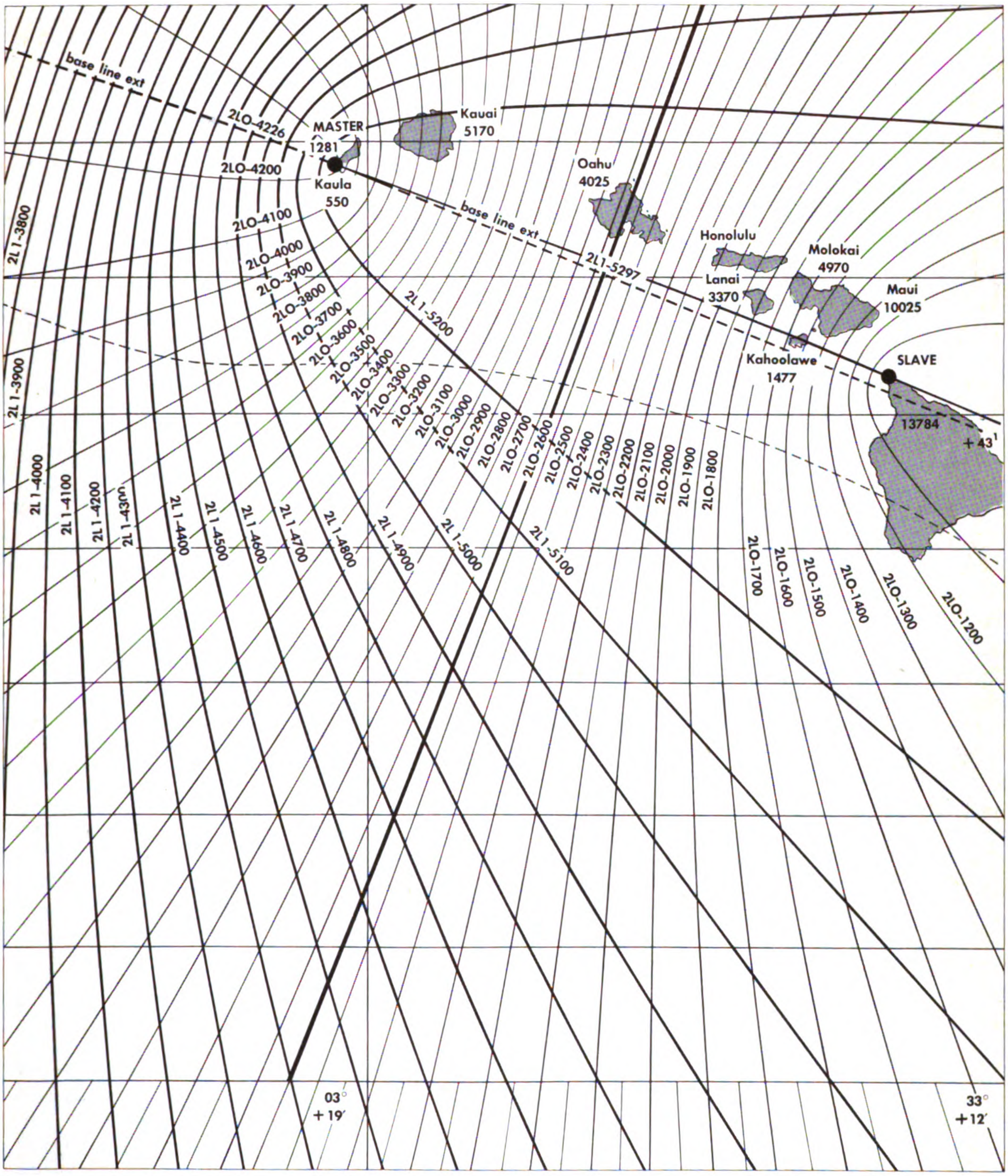
MISSILE TRAJECTORY CURVES

Missile trajectories follow many types of curves, most of them determined by the random instantaneous position of missiles relative to target. However, an exactly predicted path is the hyperbolic course that is laid out by a Loran-type system.

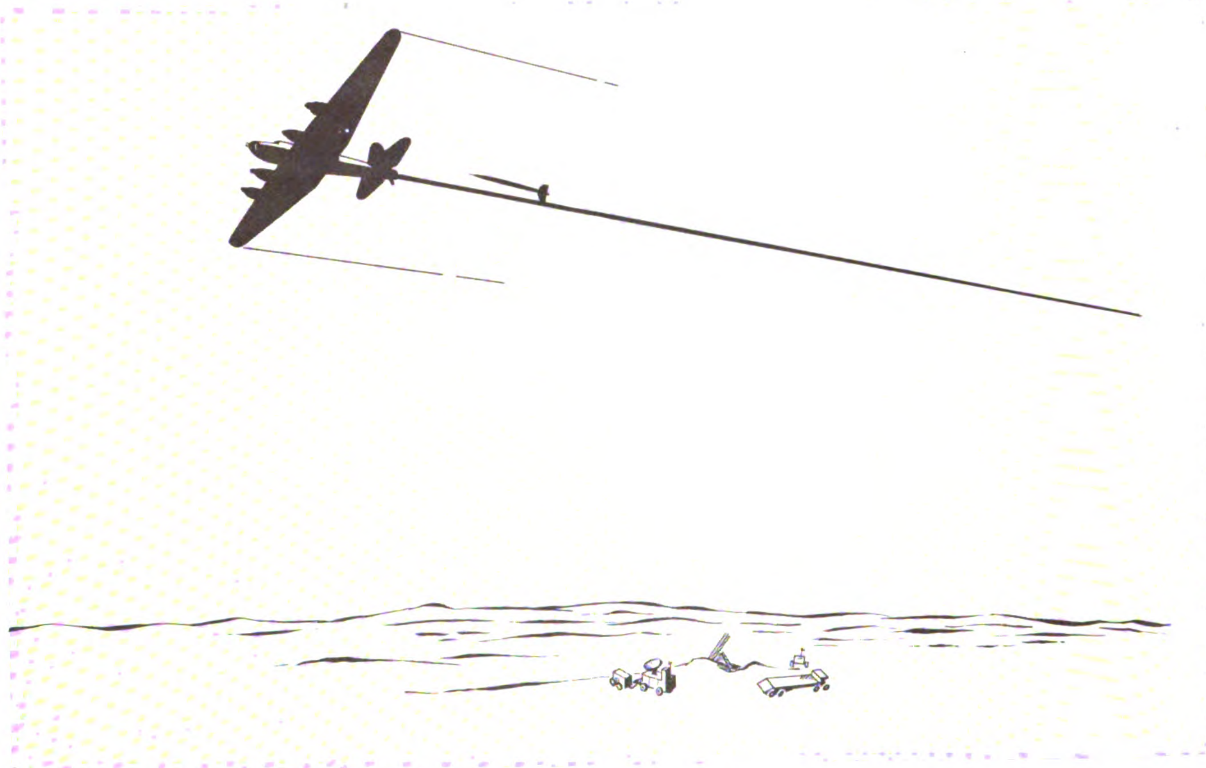
Hyperbolic System

A hyperbolic system is represented on the following page. If two radio transmitters, shown as the *master* and the *slave*, are located some distance apart and each transmit a pulse of RF energy at the same instant, a family of hyperbolic curves is generated.

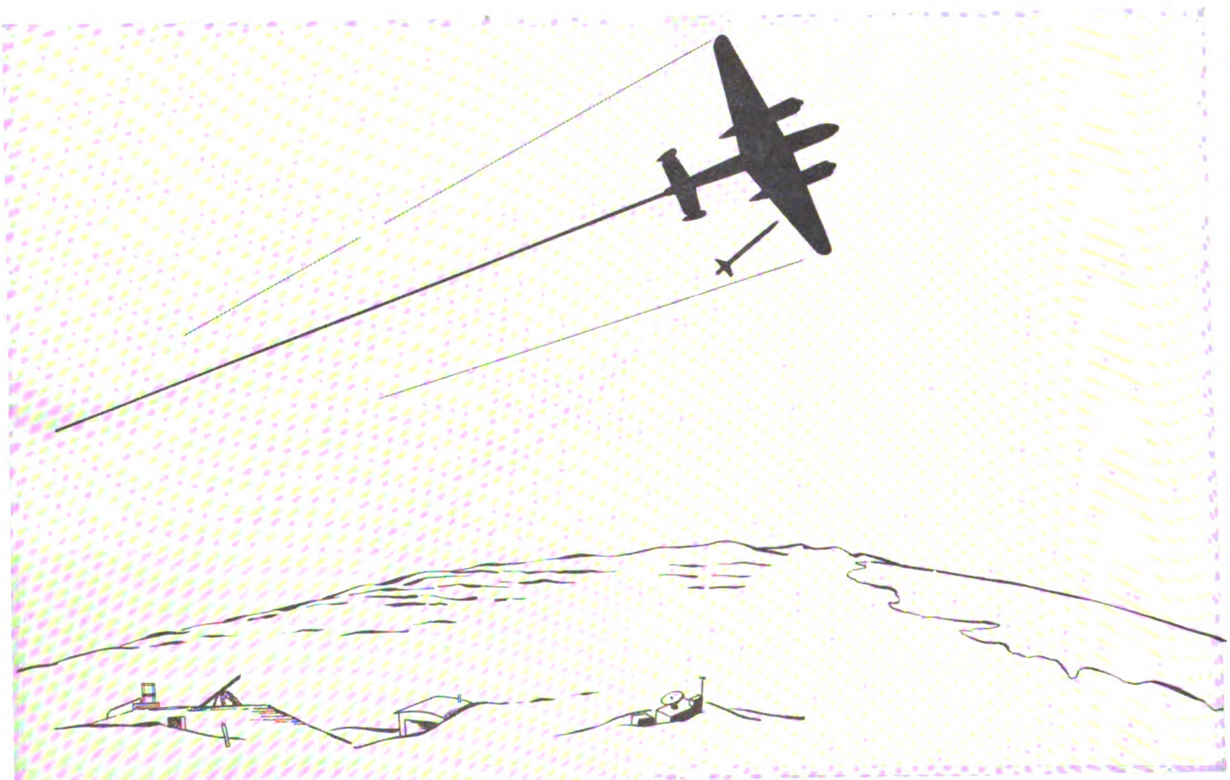
The line joining the two transmitters is the base line for the family of hyperbolic curves. About this base line the curves are symmetrical. At each point on the base line, the sequence of signals arriving from the transmitters has a different time separation. At any particular point on the base line, a singular time separation between signals is apparent. If this time separation is kept



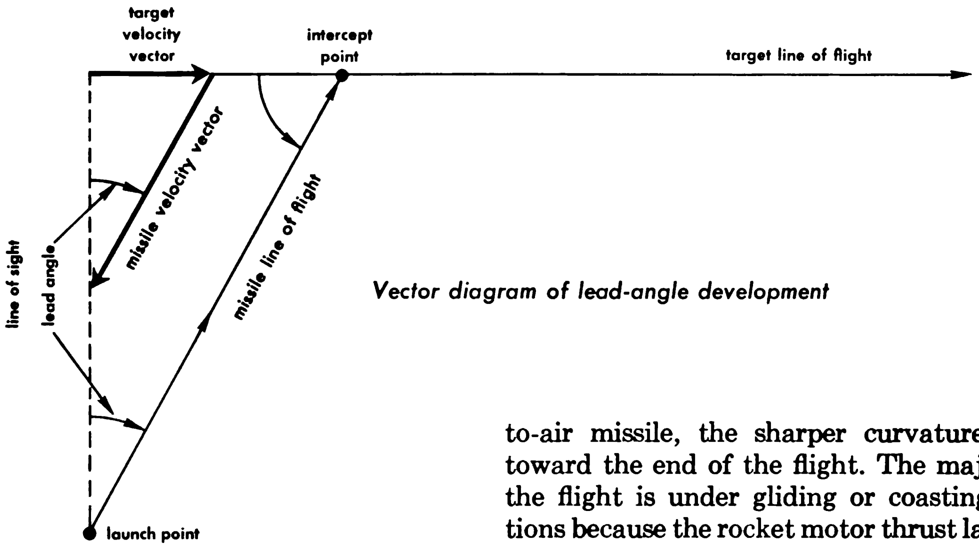
Hyperbolic lines of position generated by synchronized stations



Pursuit curve of SAM



Lead-angle interception



Vector diagram of lead-angle development

constant, movement away from the base line traces a hyperbolic curve. This is true no matter what the direction or time difference, except at the exact center of the base line. Here a locus or location of all points of the same time difference traces a straight line which is the perpendicular bisector of the base line. The hyperbolas become increasingly more curved as either end of the base line is approached from this center point. It is easily realized that the best course would be one very nearly at the center of the system because it would come nearest to being a straight course.

By adding a third station (another slave station), a second group of hyperbolas, represented in black on the diagram, are obtained. The two systems are set up to intersect at definite points from which a position is fixed or plotted. Applications of this type of coordinate or locating system are covered later in Chapter 9.

Pursuit Curve

The principle of the fixed seeker head in the homing missile has already been introduced. The seeker head is fixed to the missile. It processes the error signal in relation to its axis and the missile's axis, in as much as the missile airframe and seeker axes are common. However, such a missile must always be heading directly toward the target, and its trajectory is a pursuit curve as shown above left. Note that in the pursuit curve of the surface-to-air missile, the sharper curvature occurs toward the end of the flight. The majority of the flight is under gliding or coasting conditions because the rocket motor thrust lasts only a small portion of the flight. At the end of the flight when the greatest requirement for turning occurs, the missile has the least capability for turning. The possible turn required of the missile would quite often be beyond the aerodynamic limits of the missile.

Lead Angle Course

The drawbacks to pursuit type homing can be overcome by swiveling the seeker head in its mount so that it can look in directions other than that in which the missile is going. The trajectory flown by such a missile is known as the collision course or lead angle course. Notice in the accompanying drawing, which illustrates lead angle interception, that this course approximates a straight line, thus subjecting the missile to a minimum of maneuvering. If the angle between the seeker axis and missile axis remains fixed, the missile will intercept the target. The fact that this angle stays constant when both target and missile are in motion compensates automatically for any differences in speed between the two. The representation of such a problem can best be done by vectors where the magnitude of speed is also taken into account. The preceding diagram represents this development. A slight curve in the trajectory is caused by the deceleration of the missile after burn-out.

Beam-Rider Course

The problem with the beam-rider type of missile is equivalent to that of the homing missile. The beam rider's course is a result of the

beam movements because the beam rider stays within the radar beam controlling it. For the same reason that led to the elimination of the pursuit curve, the controlling of a missile by a target tracking radar has been given up. The tracking radar now manipulates a controlling radar so that the target position is predicted, and the missile flies a collision course to that point.

Artillery-Type Trajectory

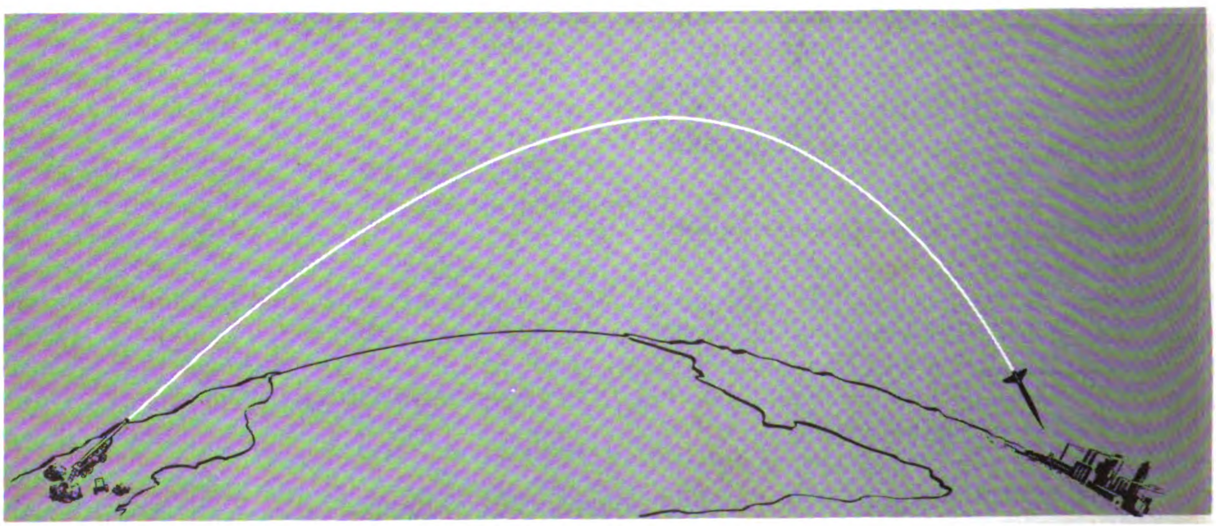
A parabolic trajectory is the classic curve accepted for a surface-to-surface missile in free flight. But now, if by extending that same free flight until the missile no longer returns to the earth but becomes a satellite traveling about the earth, the trajectory is known as an ellipse, as are all solar system orbits. Evidently then, at some point the path of a missile undergoes a change of characteristic and the transition from the parabolic to the elliptical curve takes place. However, the missile has been flying under conditions dictated by the same forces whether it falls back to earth or continues on an orbit about the earth. Therefore, except for air friction, the actual curve that the missile follows has been the same in both cases.

An erroneous assumption in the theoretical consideration of the ballistic path led to the belief that the path was a parabola. That assumption said that the earth was flat and gravity acted parallel to itself. The truth is, as you know, that the earth is a globe and gravity acts toward a center. The conditions in analytic geometry of the equation parameters that define the curve of a parabola allow but a single shape for the parabola. However, the conditions for elliptical curves include an infinite number of shapes as determined by their eccentricity. By making use of actual conditions of earth shape and gravity direction and analyzing a free fall path, it is found that the equation parameters are those of an ellipse rather than of a parabola even if only by a very small amount. Thus the curve of any ballistics trajectory is neglecting air friction and is actually elliptical.

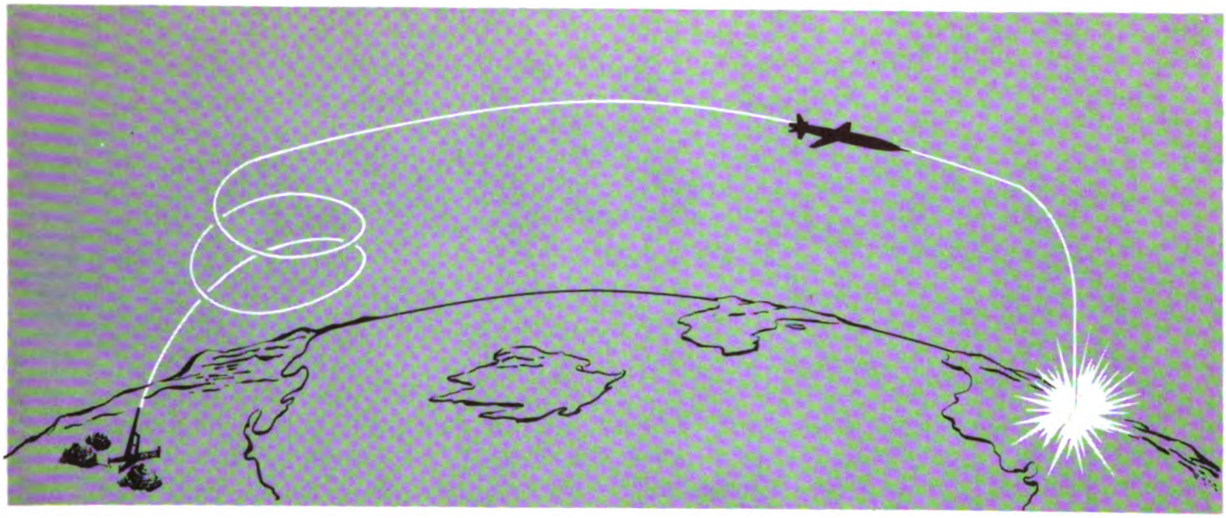
In surface-to-surface missiles, there are several possible trajectories. The choice of trajectories is made before the missile is designed. Once built, a missile is limited to flying the trajectory to which it is bound by design.

The word trajectory is commonly associated with the ballistics trajectory of a rifle or artillery piece. This artillery-type trajectory has launching and attack angles of less than 45 degrees. It follows the curve shown in the sketch below. At altitudes below 30,000 feet, this type of trajectory would be limited to subsonic flights because of the friction developed by the density of the atmosphere.

FLAT TRAJECTORY. A flat trajectory is the type employed by aircraft and surface-to-surface missile at the present writing, The mid-



Artillery-type trajectory

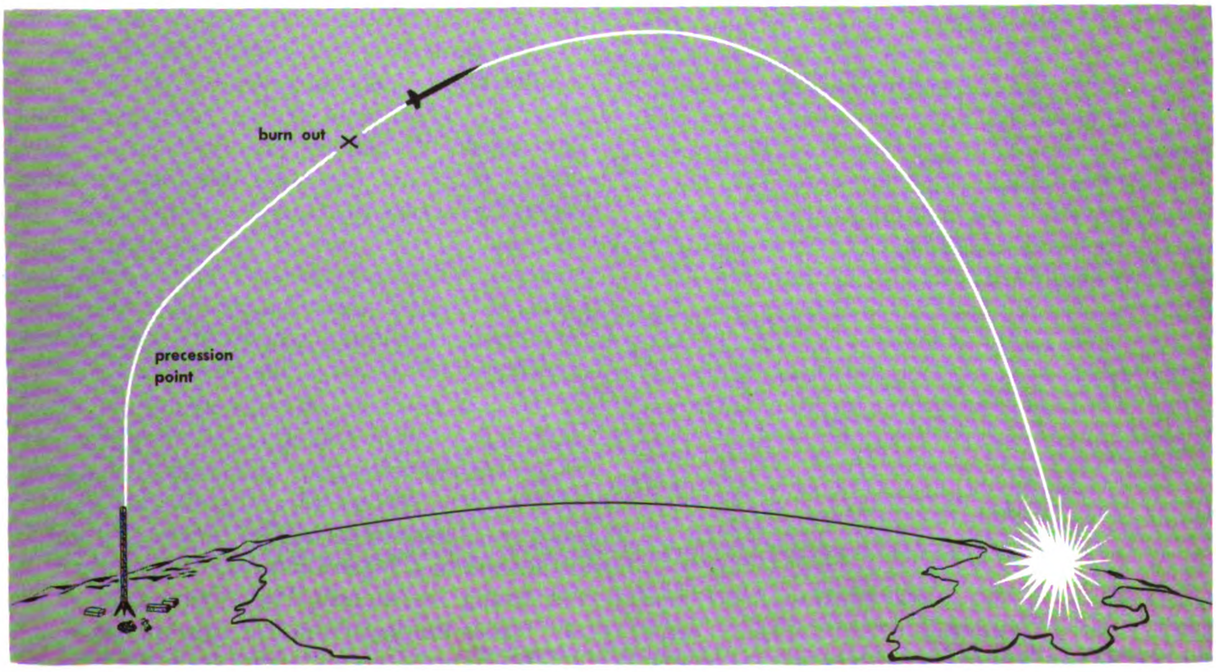


Constant-altitude trajectory in which corkscrew climb is used

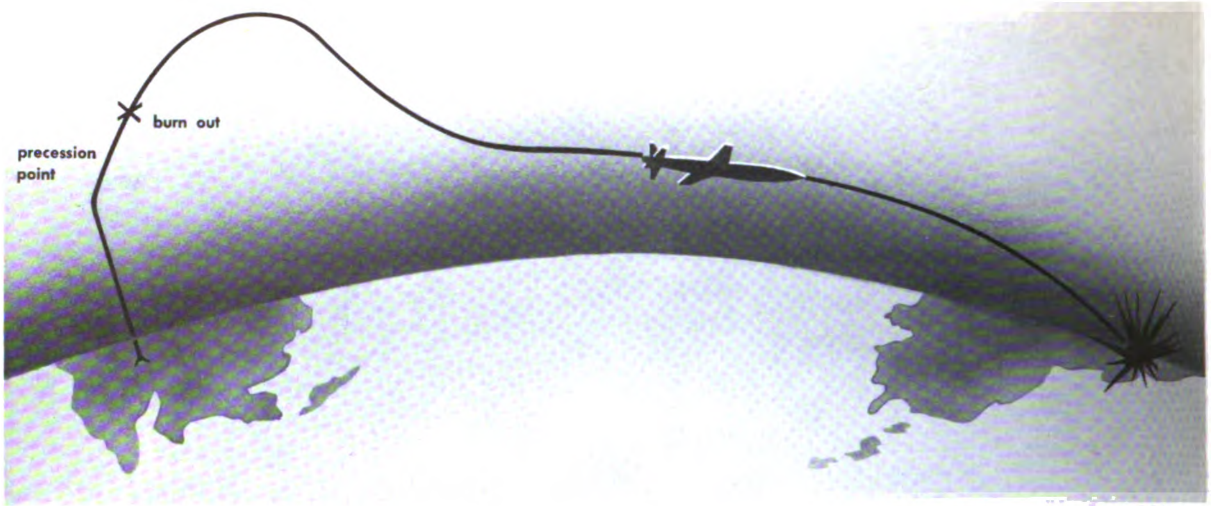
course altitude is maintained constant between 30,000 and 90,000 feet. The altitude chosen is at the level best suited to the propulsion means employed. Turbojets or ramjets can serve to drive a missile at supersonic speed in this region. A representation of such a flight is given in the preceding figure, which illustrates a constant altitude trajectory.

Rocket Trajectories

Rocket trajectory is similar to artillery trajectory. However, the difference of definition lies in the vertical launch angle of the rocket. Note in the diagram, High-Angle Rocket Trajectory, that the rocket is launched straight up until it clears the impeding portions of the atmosphere and is then precessed or tilted to an angle designed to give



High-angle rocket trajectory



Winged rocket trajectory

the proper range. The attack angle of such a flight is rather high as illustrated in the figure. Such a flight occasions cooling problems upon reentry to the earth's atmosphere.

Greater range from a rocket flight can be accomplished by adding wings to the rocket. The modified trajectory obtained under such conditions is shown as the winged rocket trajectory in the drawing above. The first portion of the rocket flight through burn-out time and over the peak of the trajectory is the same as before. In gliding down the trajectory, the winged rocket enters the portion of the atmosphere which is of sufficient density for the wings to exert a lift. This lift causes the rocket to rise again in a lift bounce as noted in the drawing. If desired, the lift bounce can be altitude controlled to the extent that a constant altitude is maintained for some distance. Since the missile is decelerating during this time and will soon lose the speed necessary to maintain the lift, it then starts a steep glide toward the earth. Such a trajectory has the merits of range extension and minimization of friction heating effects.

NATURAL FACTORS INFLUENCING TRAJECTORY

Up to now, no consideration has been made here of factors other than the missile itself which influence the trajectory. Inasmuch as missile flights are made in or through the atmosphere surrounding the earth, some thought must be given to the effects on a trajectory by meterological phenomena (the atmosphere) and terrestrial phenomena (the earth). Let's first look into the atmospheric conditions.

Influence of Meteorological Conditions

Besides being what we breathe, atmosphere manifests itself in the form of weather. The effects of weather limit our actions and even control to an extent how and where we live. Weather is probably the single big natural factor that has influenced the changes of civilization. However, weather as you consider it in the form of fog, rain, snow, hail, and sleet effectively does not exist above thirty thousand feet, so these factors do not beset the missile flight in the case of long-range missiles. *Wind* is the major manifestation of weather which affects the flight of these missiles. Winds are produced by differences in atmospheric pressure which are primarily a result of differences of temperature.

Air is a mixture of gases which are elastic and highly compressible, enveloping the earth to a depth of more than 150 miles. The density of air decreases as the altitude increases. If air over a certain section of the earth's surface is heated to a higher temperature, it expands, becoming less dense. The surrounding cooler air forces the less-dense air upward, as shown at the right.

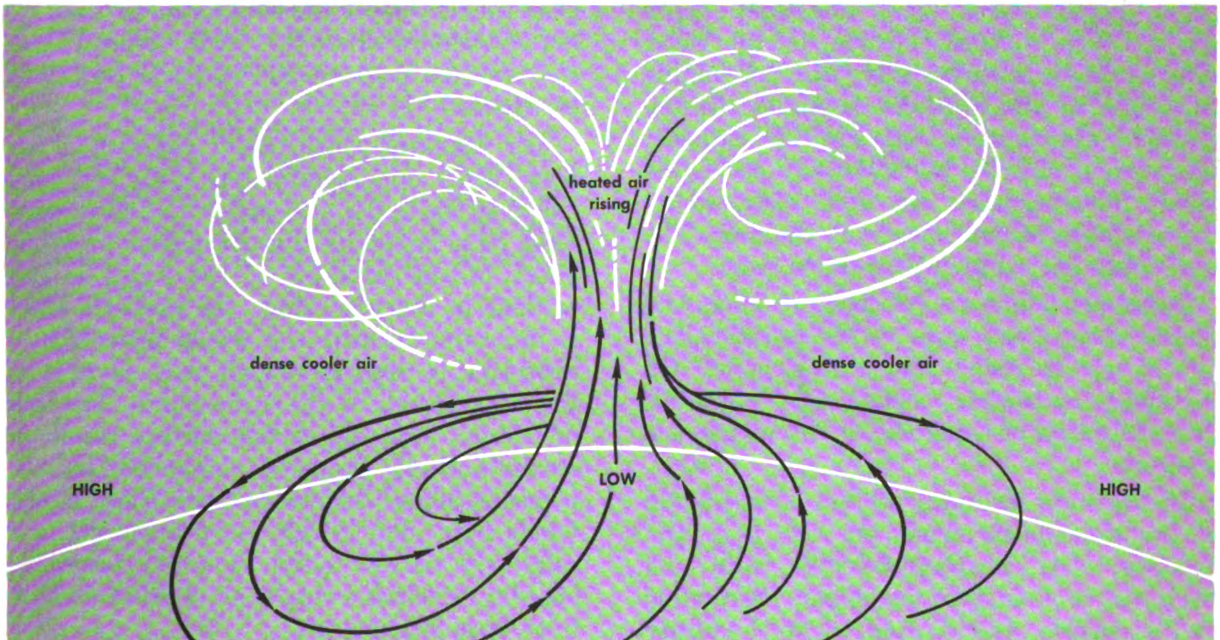
Note how the expanded air overflows to the cool air region. This overflow of air at high altitudes relieves the regions beneath it of some of the weight of air that was pressing down; consequently, the pressure beneath the heated air decreases. At the same time, the cooler regions have received the overflow of air, and the weight of air on those regions has increased. The pressure in the cooler regions increases.

These differences in the temperature of air occur because of the changing declination of the sun with summer and winter, or because some areas of the earth's surface absorb more heat from the sun during the day or lose more heat by radiation during the night than other areas, thus giving rise to temperature variations.

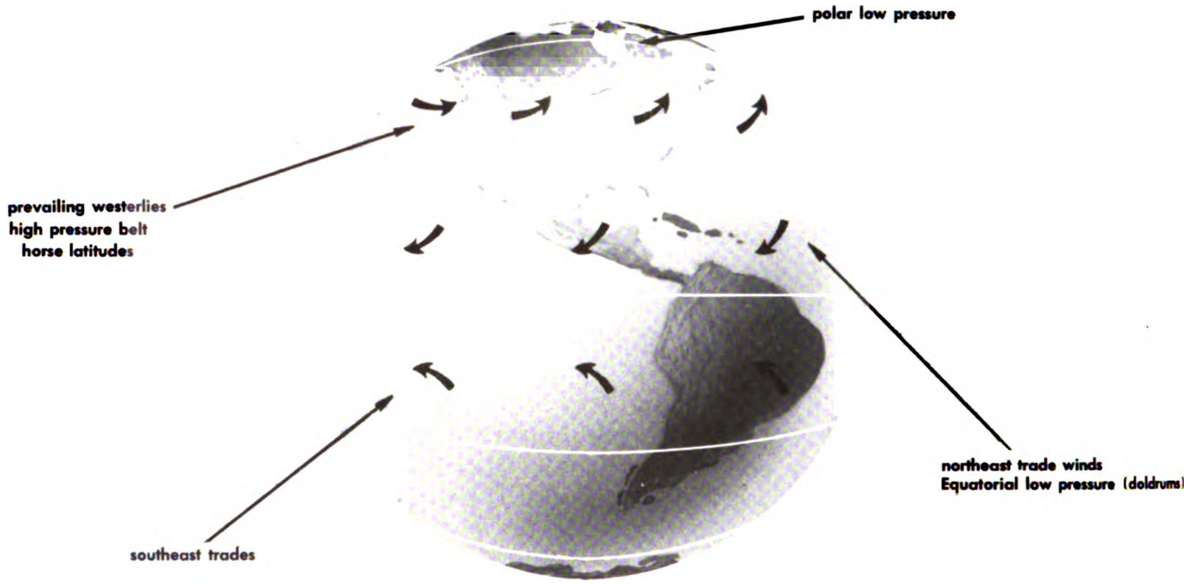
The variations lead to the flow of air in the form of winds. These winds are greatly influenced by the rotation of the earth and other factors. Because of these several influencing factors, the surface winds do not blow directly and steadily from the colder regions to the warmer regions.

WINDS. The diagram on next page shows conditions over the equator where a belt of low-pressure air surrounds the earth. On either side of this area exist belts of high pressure. The belt in the Northern Hemisphere lies mostly between the latitudes of 30° and 40° north. Beyond the belts of high pressure, the pressure diminishes toward the poles. These yearly belts are often broken up by secondary circulations of the atmosphere.

As a result of a belt of high pressure there is motion of air along the surface toward the low-pressure areas. This results in a general movement of air away from latitude 30° to 35° If the earth were stationary, the direction of this motion of air would be directly between pressure regions. However, the earth is in rapid rotation. This rotation acts to divert the air motion to its right in the Northern Hemisphere and to its left in the Southern Hemisphere. The air in motion is turned aside from its normal course down the pressure gradient, and the direction of wind at any point is at right angles to the slope of the gradient. However, within 2000 feet



Low-pressure-area formation



World wind patterns

of the earth's surface, because of friction, the wind crosses the gradient at an angle of somewhere between 45° and 90°. As a result of these conditions, the northerly winds on the equatorial side of the high-pressure belt become northeasterly (the northeast trade winds). The southerly winds on the polar side become southwesterly, the prevailing westerlies of the northern latitudes. In the Southern Hemisphere there are the southeast trade winds and the northwesterly prevailing winds. These courses are often distorted by the existence of land masses in the area. The true winds exist only in large open seas.

Doldrums, which are relatively windless and humid rainy areas, exist at the equator. The heated air is rising above this point. On becoming somewhat cooler, the rising air dumps its load of moisture, known as tropical showers.

In the *horse* latitudes (at the polar margin of the trades) the winds are also light and variable. The weather here is clear and fresh, the cool air from the heights descending to the surface, as indicated by its humidity being relatively low.

The low in the polar regions exists as a result of centrifugal force on the atmosphere. The earth's rotation would tend to throw the atmosphere into the equatorial regions were it not for gravity and the fluid-like characteristics of air. Thus, on the surface the winds flow to the north; and, consequently, at upper altitudes the winds return southerly from the North Pole and northerly from the South Pole till they meet the equatorial air over the horse latitudes where they both descend again to form the trade winds and the prevailing westerlies.

JET STREAMS. Until recent years there has been little opportunity for study of wind travel in the upper atmosphere. Wind theories must be modified for this consideration. While effect of the rotation of the earth is still evident in the wind flow, the pressure effect is opposite, because the winds blow from a low-pressure area to a high-pressure area at the upper levels of the atmosphere. The lack of friction with the earth's surface results in currents more properly controlled by the pressure areas. Jet streams are these strong winds high aloft which move eastward in a relatively narrow band at often incredible speeds. Velocities as high as 310 miles an hour have been clocked.

Horizontally, the jet streams may veer as far south as 20 degrees latitude and as far north as 70 degrees, well above the Arctic Circle. Vertically, their altitude varies from 20,000 to more than 40,000 feet. One theory

is that warm and cold currents of air collide, turning potential energy into energy of motion. Another theory is that the earth's rotation provides the necessary shove.

Prediction of these jet streams without the necessity of going up and actually measuring them is the problem at present. It is believed that four cloud types are keys to the whereabouts of jet streams. These formations include cirrus streamers, high cirrocumulus, altocumulus, and the billowing type of altocumulus clouds often extending from horizon to horizon. Also, celestial observations show that stars seem to twinkle faster when viewed through high, rapidly moving air. A scintillation count of this twinkle may provide a clue as to both the speed and altitude of these winds.

These jet streams are important in the planning of a missile flight. They will be utilized once they have been sufficiently charted so as to become predictable. Eastbound trans-Pacific flights already have been riding these jet streams during the winter when the streams are favorable. By using the jet streams, the planes are able to skip the usual Wake Island refueling stop. You can see what great advantage there would be in knowing the paths of any jet streams along the route of an intercontinental missile flight so that a favorable one could be utilized and the unfavorable ones bypassed. Air commerce of the future likely will ride the jet streams in the same manner as the old clipper ships ran before the trade winds.

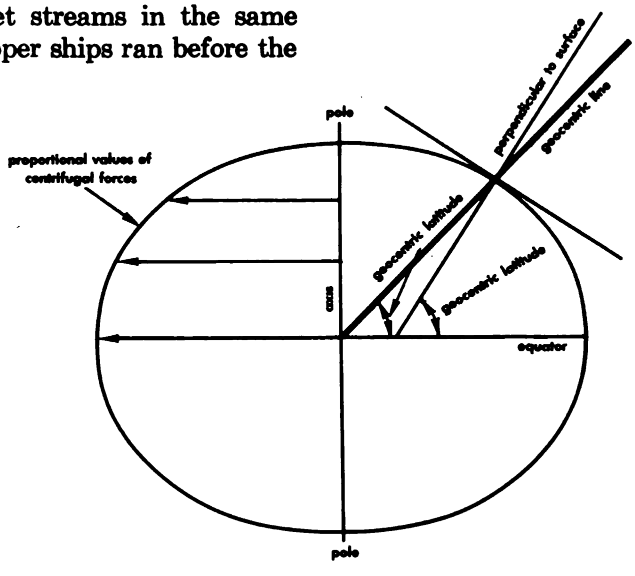
As stated, these meteorological effects are present in the flight of a missile. They manifest themselves to a great extent in the form of drift, modifying the course of a missile across the earth.

Other phenomena of the earth also can have noticeable control over the course of a missile. Some effects are felt through the guidance system alone, such as those of gravity direction, or magnetic distortions. The rotation of the earth is the cause of *coriolis force* which acts directly on a moving missile. This force must be compensated for by the guidance system.

Effects of Gravity on Missile Flight

You are sure to believe that a level resting on a surface is horizontal when the bubble of the level is centered. Just as surely, a plumb bob defines vertical in relation to the surface of the earth. Gravity is the force acting on these two items to give the results mentioned. In physics, gravity is defined as the force of attraction between two centers of mass. A little study will lead to a paradoxical situation in connection with this definition.

The earth is not a true sphere but a figure known as an ellipsoid. This is a sphere pushed in at the poles. If you study the figure below, you can see that a vertical to a tangent at any point on the surface, except the equator



Exaggerated elliptical cross-section of earth

and the poles, will not pass through the exact center of mass. It developes, then, that either the definition of gravity is wrong or that there is error in the means of measuring the direction of the force of gravity.

Actually, it is just that the definition of gravity as you note its action is incomplete. The force that causes the plumb bob to align normal to the surface of the earth is the force known as *apparent gravity*. Apparent gravity is the resultant of the true gravity plus the centrifugal force of the earth's rotation. The earth rotates about its axis. As a result, centrifugal force acts in a direction perpendicular to this axis. The amount of centrifugal force applied to an object on the earth's surface is proportional to the distance from the axis. Thus, centrifugal force is greatest at the equator and decreases to nothing over the poles as the radius to the axis decreases to zero.

Because apparent gravity is the resultant of centrifugal force and true gravity, the earth's shape adjusted itself to the elliptical. This adjustment developed as apparent gravity acted on each of the particles composing the earth.

The apparent gravity acts perpendicular to the surface of the earth except in local instances where large mountains or particularly dense underground deposits deflect the apparent gravity in the direction of the denser mass. Some deviation of gravity direction from the true vertical also exists with movements due to tides.

Except for these local variations in gravity, the force of gravity can be considered as acting perpendicular to the earth's surface on a body at *rest* on that surface. Any body in motion is contributing a further force which alters apparent gravity. Because centrifugal force contributes a substantial amount to apparent gravity, any motion on or over the surface of the earth by an object alters the effect of centrifugal force upon the object and thus alters the apparent gravity acting on that object.

Effect of Coriolis Force

A missile moving in any direction over the surface of the earth tends continually to turn toward the right in the Northern Hemisphere and toward the left in the Southern Hemisphere. Its actual motion is a resultant of this tendency and of whatever forces may act on it. This deflection toward the right or left is an effect of two motions, namely the rotational motion of the earth and the movement of the missile relative to the surface of the earth. This deflective force is known as *coriolis force.*

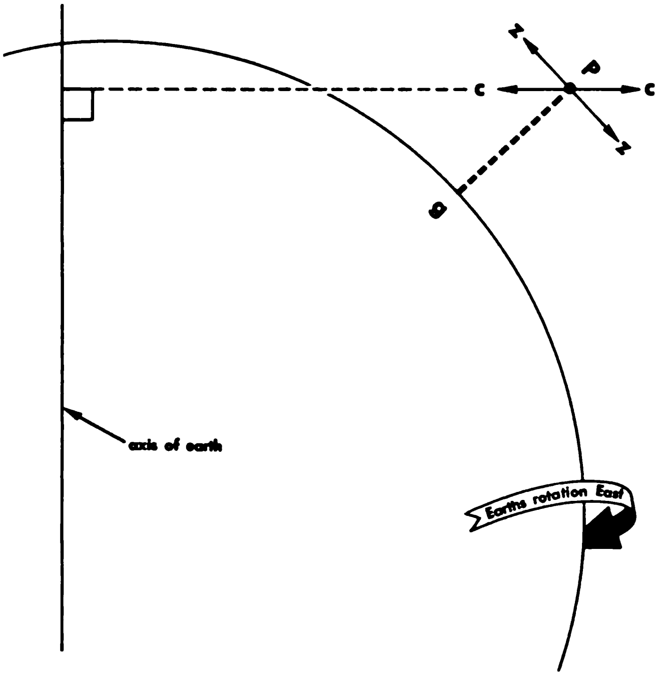
Consider an object moving toward the east in the Northern Hemisphere. The earth rotates in an easterly direction. The object experiences the effect of even greater centrifugal action since it is moving to the east. The action of this increased centrifugal force is straight out from the earth's axis; this results in a motion to the right of the eastward movement. Examine the particle in the illustration on the right in which "P" is a direction into the page around the earth's surface.

If "P" were relatively fixed over the earth's surface, the only force acting on it would be apparent gravity along the line "Pg." As "P" moves eastward faster than the surface of the earth, greater centrifugal force is exerted on it because of its increased angular velocity about the axis of the earth. The centrifugal force acts perpendicular to the earth's axis through "P" in the direction "c" away from the earth. However, the component of vector "Pc" that "P" experiences is in the horizontal direction toward the equator or to the right of its actual motion. The component is represented by "Pz."

If "P" were moving westward, the centrifugal force would be decreased in the direction of "Pc′." The horizontal component would be "Pz′," again to the right of the direction of travel.

This drifting to the right or left is referred to as *coriolis acceleration* and is tabulated for use by air navigators in the Air Almanac. It is necessary, then, to make a correction for coriolis force on any aerial flight over the earth, especially for a missile traveling at high speed.

Coriolis force can also be recognized as the force responsible for the deviation of the winds to the right in the Northern Hemisphere and to the left in the Southern Hemisphere.



Action of coriolis force

Magnetic Field

The earth is a magnetized body. It acts like a great spherical magnet with poles of unlike magnetism situated within the Arctic Circle and the Antarctic Circle. As is common with magnets, a magnetic field surrounds the earth, and it is this field which influences the magnetic compass. The earth is not magnetized symmetrically with respect to the geographical poles. The magnetic pole in each hemisphere differs in geographical position by a large and unequal amount from the geographical pole. Generally speaking, a compass needle does not point in the direction of the geographical pole, and the error varies from point to point. This angle of error is known as magnetic variation or declination.

MAGNETIC VARIATIONS. There are not only magnetic variations between places on earth, but also daily, monthly, and annual variations which are actually minor compared to the progressive variation which is constant over the centuries. Inasmuch as all these quantities are measurable, they could have great effect on calculations involving precise magnetic knowledge.

DIP. A magnetic needle freely suspended in the earth's magnetic field shows a vertical movement in aligning itself with the earth's total field. The vertical component of the earth's magnetism is known as *dip*. Dip does not exist at the magnetic equator which lies close to the geographic equator. Moving away from the magnetic equator, the dip increases until it becomes vertical at the magnetic pole.

INTENSITY. The direction in which the freely suspended magnetic needle is aligned along a line of force indicates the field. In this position it is parallel to the maximum magnetic intensity of the earth. The intensity is the number of lines of magnetic flux in a standard cross-section area. Equipment operating on the same principle as the flux valve of the gyrosyn compass can be used to measure this intensity.

MAGNETIC STORMS. The earth's magnetic field is subject to occasional fluctuations lasting from a brief period to several days. These fluctuations are called *magnetic storms*. They are due to sudden changes in the electric currents which circulate within the earth and in the region surrounding the earth. They occur apparently at random, having some correlation with sunspot activity. They may occur simultaneously over the whole earth or may be restricted to a certain region.

The range of their effect upon a compass does not often exceed half a degree in lower latitudes, but it is of greater concern in the higher latitudes. The presence of the aurora is an indication of their occurrence.

Local disturbances in magnetism occur in regions where mineral substances within the earth possess magnetic properties. The amount of this effect can be determined only by actual survey.

PRACTICAL NAVIGATION

In surface-to-surface navigation, it is necessary to begin where the usual air navigator leaves off. The practice of air navigation is an art, not an exact science. A navigator's accuracy is limited because of the conditions under which the navigator must practice this art and because of the tolerance of instruments available. With precision radio navigation aids at the terminal point for the flight and the exact placement from pilotage, the dead reckoning and celestial navigation need be accurate to within only a few miles. But a

missile seldom has radio beacon homing equipment, and it has no pilotage reference that it can be positive of, so its unattended navigation system must have pinpoint accuracy.

Target Location on Maps

The location of the target is the primary problem. The target must be accurately placed in respect to the launching point. To do this requires reference to a map. Maps are published for information as to the character of the earth's surface and location of man-made features in a form that is convenient to handle.

To make a map convenient, it is necessary to lose a certain amount of accuracy. Certain smaller features must be portrayed larger than scale if they are to even show in a scale reduction of an area. For instance, a line representing a road may be almost a mile wide according to the scale. The earth has a curved surface; and, in order to lie flat, a map must distort to a certain extent the area it represents. The way in which the surface is "stretched" to make a flat map is referred to as the *projection*. Another basic difficulty is the inaccuracy of the original information from which the map is made. This is not apparent to those who are acquainted only with the maps of this country, but other areas of the world are mapped with unknown errors existing. Even though the earth seems most stable, its surface is quite flexible. Earthquakes, tides, and erosion are all indicative of the continual shifting of the surface.

Radio Propagation

A great difficulty in exact placement of a point exists for those guidance systems using radio waves. The propagation of radio waves is quite affected by conditions of the earth over which they travel and the condition of the ether through which they travel. Air has an effect because of several factors. You are familiar with the ionosphere's skip effect on shorter waves as illustrated at above right.

The havoc that magnetic storms can wreak on radio reception also is well known. As the frequency of the radio waves is increased to VHF and beyond, even the density of the air has an effect on transmissions.

You will remember that in the discussion of weather, the density of the air was measured in terms of the pressure it applied. The refraction of waves passing through media of different densities was made clear in physical principles. Their refraction also occurs in radio waves in the presence of varying air densities.

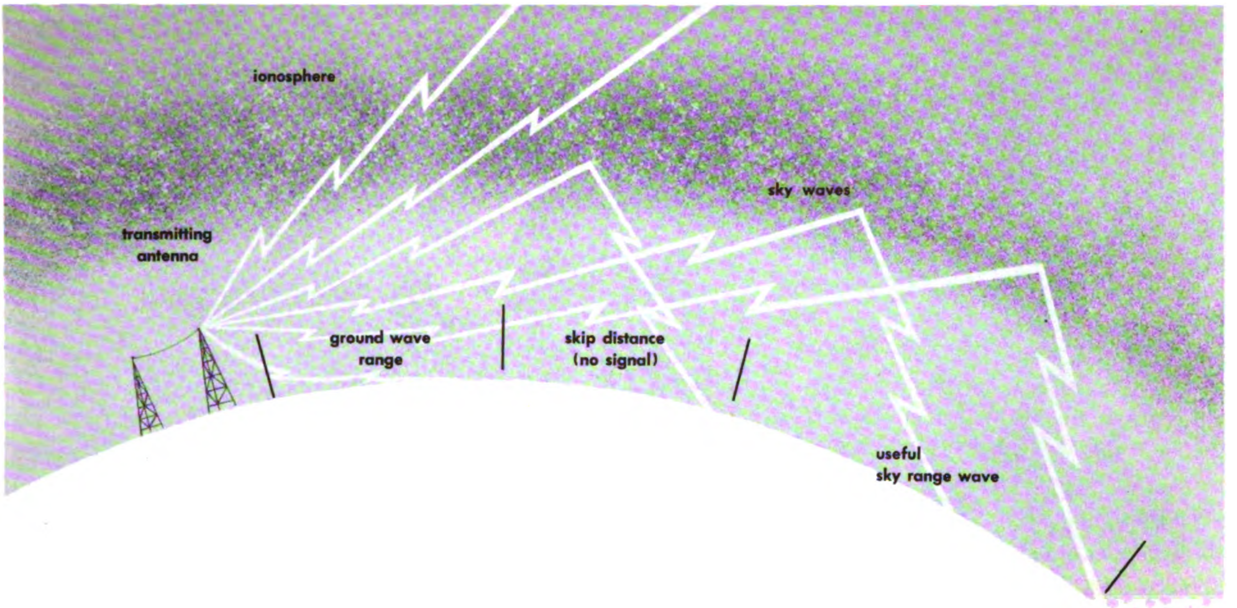
The condition of temperature inversion, when warm air overlays cold air, is desired by amateur radio operators for purposes of long-range communication on VHF bands. Note in the illustration below right that warm air, being less dense, refracts the radio waves back toward the cooler air which is at the earth's surface. There are conditions, then, in which radio waves *do not* travel along a line-of-sight route and thus cannot be depended on to do so.

Line of Position (LOP)

If you are on a line which has a definite bearing to a known point, you are on a line of position (LOP). A line of position is the plot of all positions in the line of the particular bearing. LOP may be derived from a compass or radio bearing, Loran sights, or celestial sights. The position itself is not known until another LOP is obtained which intersects with the first LOP. The point at which the two LOP's intersect is known as a fix. The line traced by a series of fixes is called a course or track. An accurate fix is also obtained when the course crosses a landmark which is outstanding.

Course Terms

In order to maintain a course, the missile track must be headed along that course. The heading is most often taken from a magnetic reference, in which case it is known as magnetic heading. Because of magnetic variations, a true heading is often used; this gives the actual geographic direction of flight. However, a true heading may not result in a course of the proper direction due to drift. The greatest drift occurs because of crosswinds through which the missile is flying, as shown by the diagram on page 420. Some drift may also be the result of the missile being out of lateral trim. A true heading is usually chosen that is correct for the course between two check points if on a long flight, or between the departure point and target point on a shorter

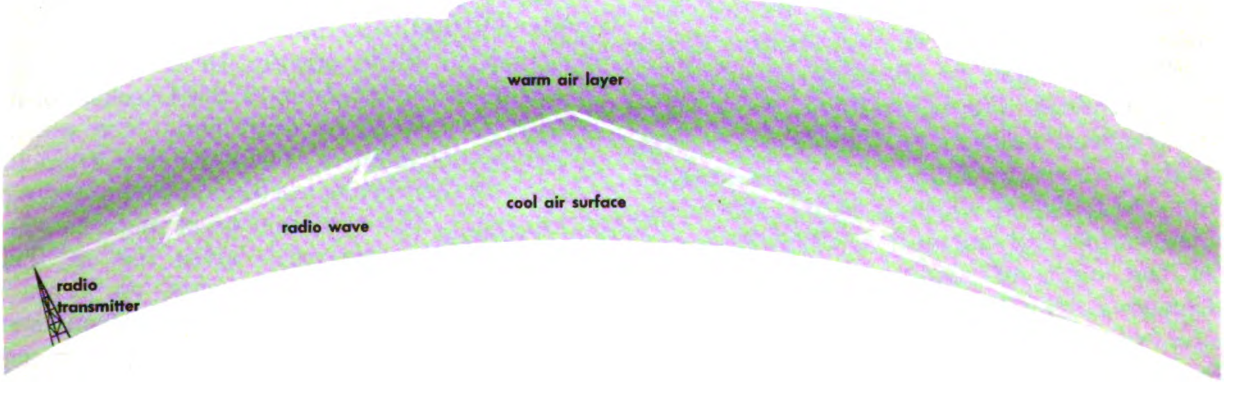


Ionosphere effects on radio transmission

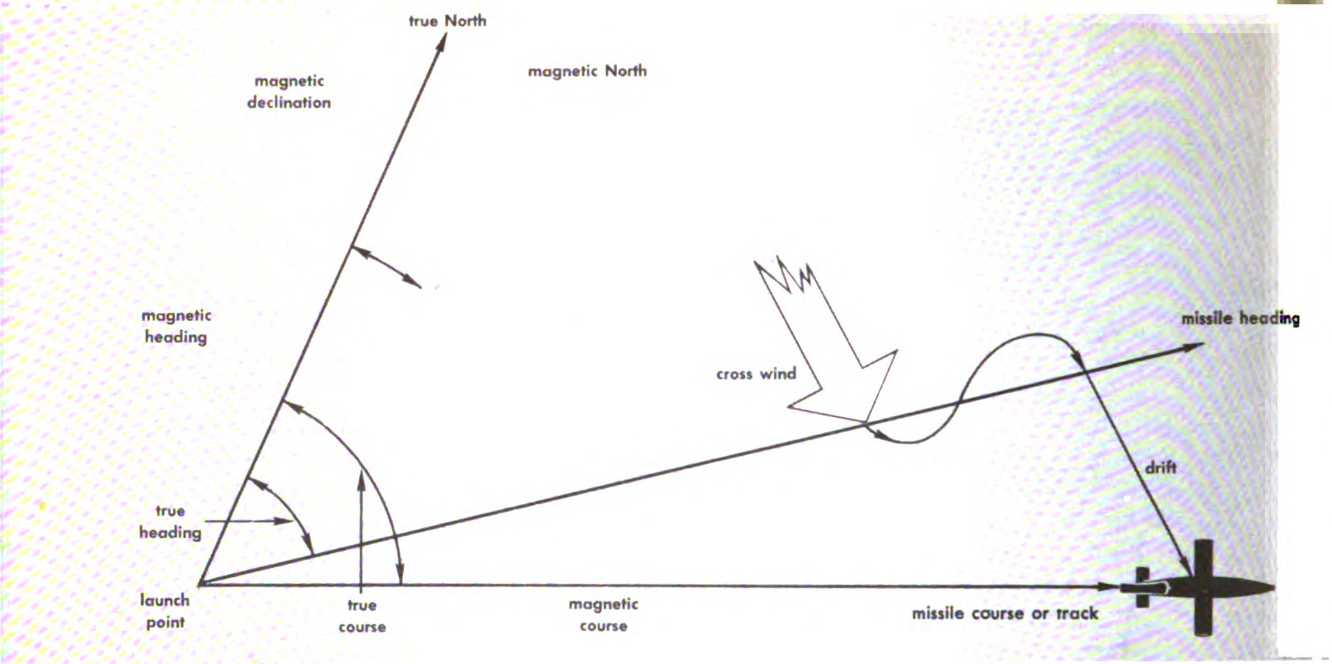
flight. The true-heading course is called a *rhumb line*. A rhumb line is a straight line which is *not* the shortest distance between two points on the surface of the earth. A rhumb line possesses the same angle with all the meridians of longitude which it crosses and is therefore a single-heading course.

Great Circle

The shortest distance between two points on the surface of the earth is measured along a great-circle path. The illustration on page 421 shows the meaning of a *great circle*. The great circle connects two points on a line that a flat plane would make on the earth's surface if it passed through the two points and the center of the earth. The equator is a great circle; so are the meridians (lines of longitude). All the parallels of latitude other than the equator are *small circles*. The great circle, as you can see, is the one whose circumference is the greatest for the globe which it encompasses; thus, it has the least amount of bending and thereby approximates the shortest distance between two points. To be exact, the great circle is slightly elliptical following the shape of the earth. However, common usage has decreed that it be known as a *true great circle*.



Refraction of radio wave by temperature inversion



Drift caused by crosswinds

CELESTIAL NAVIGATION

In order to simplify the conception of the heavens for use in celestial navigation, it is assumed that the stars are on the inner surface of an infinitely large sphere which rotates about the earth in a manner pictured in the sketch on page 422. This conception is not exclusively modern, having been proposed by Aristotle and his followers several hundred years before Christ. The stars all rotate together, but the sun, moon, and planets have eccentric orbits, and are treated separately. The celestial sphere is assumed to, and apparently does, rotate on an axis that is the extension of the earth's axis. At any instant, a point or star on the celestial sphere is over a corresponding point on the earth because this geocentric system consists of two concentric spheres. Through the use of trigonometry and a sight angle measuring device such as a sextant, an observer's position can be determined. The navigator knows a star's geographical position on the earth and uses that as the center from which to measure an arc, the radius of which is ascertained by his observation and calculation. This arc is part of the circle of equal altitude, so called because at all points on this circumference the star will have the same angle above the horizon at that instant. The circle of equal altitude is a celestial LOP. The intersection of two or more celestial LOP's gives a position or fix.

Stars as Navigational References

Of all the celestial bodies stars give the least amount of light, but nevertheless they are the most desirable navigation references. One reason they are desirable is that they are practically point-sources of light and can be accurately sighted on. An even more desirable characteristic of the stars is that they are at such vast distances from the earth that the light rays coming from them are essentially parallel. This is not true of the divergent light from the celestial bodies in our solar system.

PLANETS. To the observer on earth there is no relative motion among the stars, while there is considerable motion among the planets, sun, and moon. The motion of the planets is quite erratic. In fact, the word *planet* meant *wanderer* to the Greeks. As the earth rotates in its orbit about the sun, so also do the planets. Their distance from the sun and their orbits are different from those of the earth. This combination of motions when viewed from the earth gives rise to their apparently erratic movements. At times a planet will even be seen to move from west to east. For these reasons, planets are not used for celestial navigation.

On the other hand, for purposes of navigation, stars are considered as fixed to their position on the celestial sphere. Actually, there is a minute detectable motion among the stars, but it will take many years before this motion has an effect on ordinary navigation.

MAGNITUDE SCALE. Stars are classed on a brightness scale. This scale is known as *magnitude*, and the lower the number the brighter the star. Those stars visible to the naked eye on a clear night are 6th magnitude and brighter. Ordinarily, only stars of the first two magnitudes are used for navigation.

Stars also give off light of different colors (wavelengths), an important fact when using an electronic device to track them. Certain phototubes are more sensitive to light of a particular color. In such an instance, color-corrected magnitudes are used. This requires a check and relisting of the stars according to the amount of light of certain colors or wavelengths emitted.

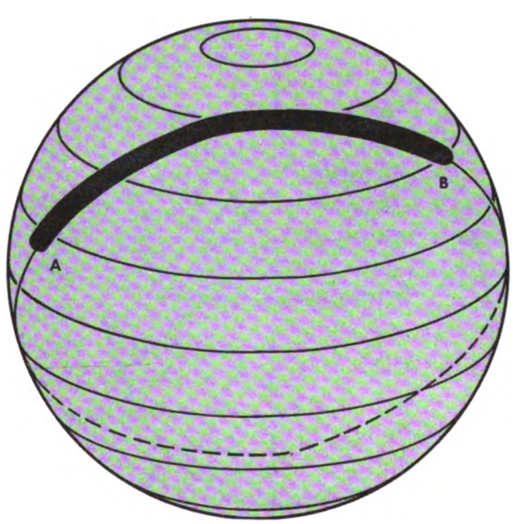
Celestial Coordinates

The celestial sphere has certain features corresponding to those of the earth. However, because the earth rotates within the celestial sphere, latitude and longitude cannot be used to determine the position of a star. The celestial sphere uses declination and sidereal hour angle for locating coordinates. Declination is the celestial latitude in respect to the equinoctial or celestial equator. The sidereal hour angle is measured westward from an arbitrary meridian known as the *first point of Aries*.

The first point of Aries is the point at which the sun crosses the equinoctial line at the time of the vernal equinox and corresponds to the prime meridian of earth as a starting point for longitude. The sun's path on the celestial sphere does not follow the equinoctial line but a curve called the ecliptic, because the earth is tilted in respect to its orbit about the sun. The ecliptic reaches its northernmost point at the summer solstice and crosses the equinoctial line during the equinoxes. Thus, the sun's declination varies from north 23°27′ to south 23°27′ during the course of a year.

Latitudes

Certain small discrepancies arise when the ellipticity of the earth is considered. It is possible to develop three different latitudes for the same spot because the earth is not a perfect sphere. These are the geocentric, astronomic, and geodetic latitudes. Referring back to the illustration showing the elliptical cross-section of the earth, on page 415, you see that the geocentric latitude is the angle which a line through a point to the center of the earth would make with the equatorial plane. The astronomical latitude is the angle



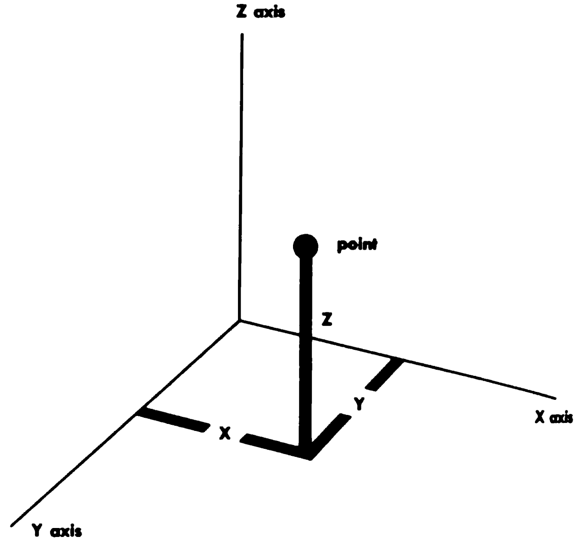
Plane of great-circle route from A to B

with the equatorial plane that a line perpendicular to the earth's surface would make. Geodetic latitude is the same as the astronomical latitude except for corrections in the direction of gravity due to local conditions.

The ellipticity is magnified in the illustration. Actually the ellipticity of the earth is small enough so that at latitude 45° there is only 11 minutes difference between the two latitudes, geodetic and astronomic. But even this difference, if uncompensated for, could have considerable effect on the accuracy of a long missile flight.

As was stated earlier, the greatest problem is predicting the location of the target point at the time the missile will arrive so that a hit can be effected. This problem is obvious in the intercept missile which is fired at a moving aircraft. The aircraft is not only capable of motion in its initial direction but may also take evasive action. The intercepting missile must have superior flight characteristics or a superior computing facility so that it can outwit the maneuvers of its target and complete the interception.

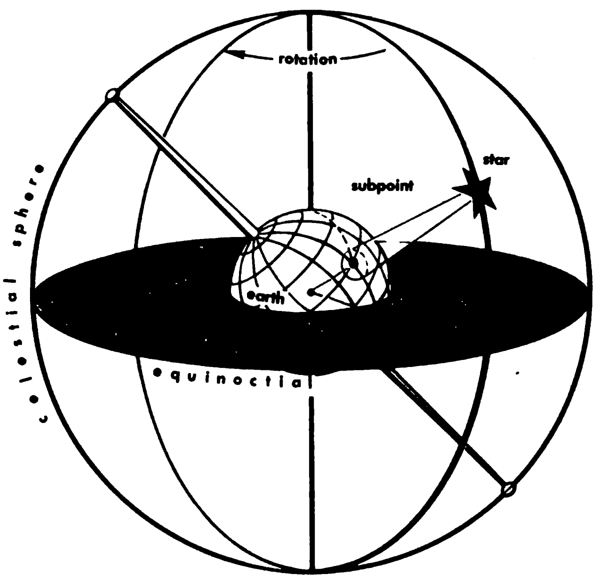
Although it is not as obvious, the problem of a long-range missile flying toward a surface target is much the same, if not more complex. Recall that the target is on the earth which



Rectangular coordinates

is rotating on its axis while revolving around the sun. The target then is actually moving with reference to inertial space and describes a complicated trajectory through space.

Actually, ***the missile does not fly toward the target but to a point in space at which it is predicted the target will be at the particular instant the missile arrives***. The target, being fixed to the surface of the earth, cannot



Celestial concepts

perform evasive action, but the many forces acting on the missile accomplish the same effect. The effect is sufficient to cause a miss because of the great distance the missile has to fly. A very minute uncalculated force is sufficient to deflect the missile many miles from the target on a long flight. The winds, coriolis force, and equipment tolerance, in addition to the difficulty of knowing the exact location of the target point on the surface of the earth, enhance the odds that the missile will not be able to strike the target.

The problem of a missile and its path to the target is solved in an application of mathematics termed *trajectory analysis*. The positions of the missile and target are located as precisely as possible, and then the path is charted for the missile to follow.

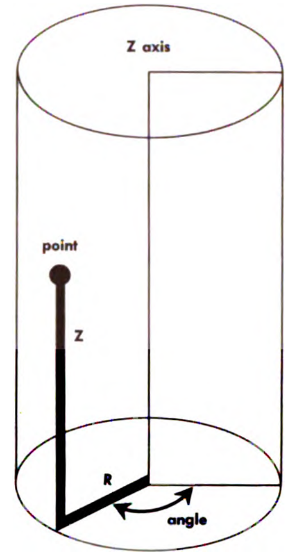
GEOMETRIC COORDINATE SYSTEMS

Because a missile flies in three-dimensional space — length, breadth, and depth — three-dimensional coordinate systems must be used. A coordinate system is familiar to you under several names: for instance, latitude and longitude are coordinates.

A coordinate system uses a reference point as its center. Generally, this point is the spot where you are, or the place from which the missile will take off. In practice, this reference point must be accurately located and predictable, for it is the source of all measurements. The measurements from the reference point or *origin* to the target are made in several fashions. The method of making the measurements determines the title of the coordinate system. Of the great variety of systems, you will need to understand the rectangular or orthogonal, the cylindrical, and the spherical systems.

Rectangular Coordinates

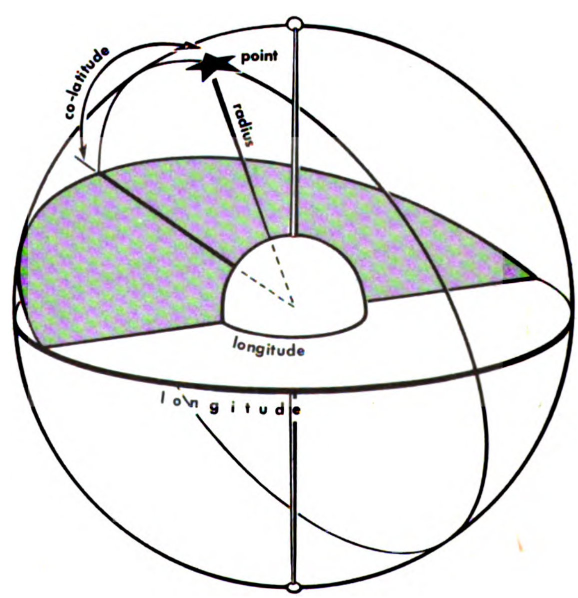
The names for most coordinate systems are self-descriptive, as is the term *rectangular-coordinates system*. This system consists of three axes which are mutually perpendicular to each other at a common point shown in the illustration above left. This point of intersection is the origin of the system. A point in space is fixed by its three distances along x, y, and z from the planes formed by each pair of lines.



Polar coordinates

Cylindrical Coordinates

In a cylindrical system the position of a point is determined by polar coordinates in a horizontal plane and the distance along the vertical axis as shown in the illustration above. Polar coordinates are not difficult to



Spherical coordinates

understand when they are used in regard to speaking of a location as being so many miles NE of one's position. Polar coordinates are a direction (angle) and a distance (radius) from the origin.

Spherical Coordinates

A spherical system is better known as the geographic system of coordinates. Note that in the previous drawing the positive location of a point is given by its distance from the origin (the radius), an angle in the horizontal plane known as longitude, and the angle from the vertical axis called co-latitude. In the sphere, the origin lies at the intersection of the axis with the horizontal plane.

Coordinate Applications

In the analytic conception of a trajectory problem, the target is often considered the origin of a coordinate system. The missile is then a point in space; the object is to bring the coordinate values to zero, thus causing a collision between the missile and the target.

However, in the case of a long-range missile, the most predictable point is usually the center of the earth. The target and missile are both represented in terms of this coordinate reference or possibly some other hypothetical point — for instance, a point fixed in space at the predicted intercept point. The equations of motion are worked out, making use of all terms which define any force affecting the motion of either the missile, the target, or both. These terms are then transformed, often several times, to give the equation parameters in values that the missile guidance and control system can effectively use.

This chapter did not attempt to present a complete study of missile trajectory. The intent was to familiarize you with the *principles involved* in missile trajectory and to acquaint you with terms you will often come in contact with.

Chapter 9

Guided Missile Guidance Systems

Many factors govern the choice and employment of guidance systems in guided missiles, but the primary condition which determines guidance system adaptation is the range involved in the flight of the missile. This condition makes logical the division of guidance systems into short- and long-range systems. Factors to consider in selecting a guidance system for use in a missile are accuracy, proof against countermeasures, distance over which the system is usable, dependability and durability, physical size of the system, simplicity of operation and adjustment, and economy in manufacture.

In the past emphasis was too often placed on the quantity of projectiles a given firepower could deliver. As a result, accuracy suffered. Artillery uses several rounds to bracket a target, correcting until a hit is made. With the advent of missiles comes a necessity for a "one shot" barrage per target. The expense and complexity of a missile led to the requirement for absolute accuracy. The first shot must be right on center. If it is not, the element of surprise is lost and an expensive missile is wasted. This means that not only must the guidance be accurate but all the prelaunch planning, calculations, and adjustments also must be absolutely correct. No tolerance from calculated requirements is permissible. Accurate fire in volume is acceptable, but volume of fire as a substitute for accuracy must be rejected.

An ideal guidance in all classes of guidance systems is one for which there can be no countermeasures. Prior to the countermeasure must come detection. A missile that emanates a radiation, or uses one generated by the launching vehicle, is detectable at the target at ranges equal to or greater than that at which the missile itself can see the target. The target can then take measures to evade or neutralize its attacker. A missile that uses a signal or radiation generated by a friendly source for guidance overcomes countermeasures by using extremely high speed, or by

short time of flight. The target defense then does not have time to react after it detects the missile.

The distance over which a guidance system can function dictates the use to which it is put. A system of limited range is, of course, best used for close targets, as in the case of aircraft interception and the like. The longer range systems are utilized as offensive bombers.

The field of view of a system is important. If the missile has a large error at the beginning of a flight and can still determine and strike its target, it is more valuable than a missile which is limited to a narrow zone of approach.

Physical and electrical factors become important in the tactical movement of a missile. Its size, the need for minimum maintenance and replacement, ease of storage and operation, and economical manufacture are all weighed in the designing of new equipment.

As the field of missile development progresses, the concept of range keeps pace. At the start, short range was classified as that which was within visual limits. At the present time a short range surface-to-surface missile is one that flies less than five hundred miles. The definition of long range has gradually expanded from that of several hundred miles to the global idea of six thousand miles or so.

In the next few chapters, some systems are covered for the historical background only. As systems evolve, the old ones become obsolete. Even so, the parent systems are worthy of consideration in the interest of understanding the reasons for trends in present circuits.

This does not mean that equipment will be increasingly complicated as time goes on. By combining principles and making accurate components, certain refinements in present circuits will no longer be necessary. These refinements only exist to compensate for possible signal errors in the first place. In the end an uncomplicated, accurate system is hoped for and anticipated.

Guidance systems could be classified according to the type of missile with which they are to be used. However, this would lead to duplication in the coverage of systems. The guidance unit in an air-to-air missile could also be utilized as the terminal guidance in a surface-to-surface missile. Certain guidance systems, such as radio command, are primarily used in the research and development phase to check on flight characteristics of the missile. Other systems, which also employ a radio link between missile and controller, are used in testing and instrumentation.

The use of a combination of guidance systems in a missile makes possible the division of a missile flight into phases. In the case of a missile that contains more than one type of guidance equipment, or in which the guidance equipment operates only a portion of the flight, the flight can be divided into phases which are defined by the type of guidance equipment operating.

A standard flight breakdown into phases which would fit all types of situations is not possible. Some missiles have a flight consisting of but one or two phases, while other more complex systems have a flight consisting of four phases, with some of the phases subdivided into periods.

All missiles have an initial or *launch phase*. During this phase the missile becomes airborne. Ordinarily, because of the acceleration due to propulsion thrust during launching, no guidance system is working. Reasons for this are covered later. In specially designed systems the guidance equipment can operate during the launch phase. The launch phase is then combined with a later phase of flight.

In the more complex systems a *calibration phase* follows the initial phase of flight. During the calibration phase the midcourse guidance system is supervised by an external system and corrected to the proper trajectory. The calibration phase lasts until the missile system is aligned, with the minimum error, to the desired trajectory.

The *mid-course guidance phase* follows the initial phase of flight in most medium- and long-range systems. Mid-course guidance is the primary system in these missiles because it must guide the missile through the major portion of its operation. Of the total flight time of the missile, mid-course guidance operates much more than all the other systems combined. It must be accurate because of the great distance over which the mid-course

guidance functions. As a result of the required guidance accuracy, the mid-course system contains the greatest refinements in circuitry and components and is therefore the most difficult to study and understand. At periods when mid-course guidance is not able to function, there are provisions to replace the guidance signal with a standby system.

Terminal guidance is the final phase of controlling a missile's flight. Terminal guidance is the phase in which the missile makes contact with its target. A short-range guidance system that allows the missile to home on its target with minimum error is used. It may also be that the terminal guidance is an inertial system whose accuracy is dependent upon the accuracy of the mid-course system. The information available to a terminal inertial system is the error information supplied by the mid-course guidance, up to the time when the terminal system took over.

The division of phases of a flight into periods is somewhat more arbitrary than the division of the flight into phases. This is true because of different requirements and principles of operation of various guidance systems. Even the separation of a flight into phases, although convenient, is no hard and fast rule. Allowances must be made for the ingenuity of guidance system design in which the functions, as outlined by the several phases, may all be combined into a common time or phase of operation.

Chapter 9 • Section 1

Short-range Guidance Systems

Short-range guidance systems can be divided into two general types: preset and command. A self-contained system of guidance falls into the category of a preset system. The intelligence necessary to the course and termination of the flight is all set into the missile prior to its launching. The usual definition of preset guidance includes only a simple system that defines heading, altitude, time or length of flight, and programmed turns.

A command guidance system is one by which the missile receives correction signals from an external source. A correction signal is a command that activates the controls a definite amount. Do not confuse it with an error signal which is the detected discrepancy from the required course, altitude, or speed. Ordinarily, the guidance system detects and develops the missile position error signal into a correction signal. After the correction signal is properly formed, it is fed into the control system.

For security reasons, a detailed discussion of short-range guidance systems originally had been omitted from this manual. Now declassified, this information on short-range systems can be found in the appendix of this manual (page 573).

Long-range Guidance Systems

In this section we'll discuss guidance methods as they pertain to inertial guidance systems, celestial navigation systems, hyperbolic systems, terrestrial reference systems, and magnetic guidance systems.

The simplest principle for guidance is the law of inertia. In aiming and firing a rifle, you attempt to give the bullet a trajectory which will terminate at the desired target. In other words, you impart an impetus to the bullet which causes it to travel along the proper path to a target. However, once you fire the rifle, you can have no further influence over the flight of the bullet. If you aim incorrectly, if the target moves, or if a crosswind comes up, your bullet will miss the target. A bullet cannot correct its course in flight or home on its target.

In a simple system of long-range guidance, a missile could be aimed and fired in about the same manner as a gun, and the flight path would be a ballistics trajectory. Direction and distance of missile flight are the items which would have to be determined accurately. If the target were 500 miles northeast of the launcher, a missile that would fly for 500 miles on the proper course until the target is attained would be launched. This would require an accurate and true direction meter and distance meter.

However, before contemplating a missile flight, exacting and comprehensive calculations must be made as to the precise flight path. The equation for this problem may contain factors due to movement of the missile about the three axes — pitch, roll, and yaw. And the equation may contain factors resulting from acceleration due to forces external to the missile and the inertial forces of the missile itself.

Beyond the limits of the atmosphere, it is relatively easy to predict the path of a body. Astronomers have been doing so for years with extremely successful results. A long-range ballistics-type missile could be considered as another body in space, and its path calculated accordingly. This would be the most difficult type of missile to intercept. The interception problem would be analogous to the interception of a meteorite.

Bearing in mind that invulnerability to enemy countermeasures is a prime requisite of military missiles, the supersonic ballistics-type missile would be quite ideal. As a matter of fact, some accuracy may be sacrificed in favor of a system which is practically without countermeasures. In the ultimate perfection of any system however, the accuracy would not suffer and error at target would be nil. Let's now consider inertial guidance systems.

INERTIAL GUIDANCE SYSTEMS

A true inertial guidance system would include, as previously mentioned, a direction- and distance-measuring system. There are several methods of accomplishing this, plus refinements to minimize errors that develop.

Basic Inertial Systems

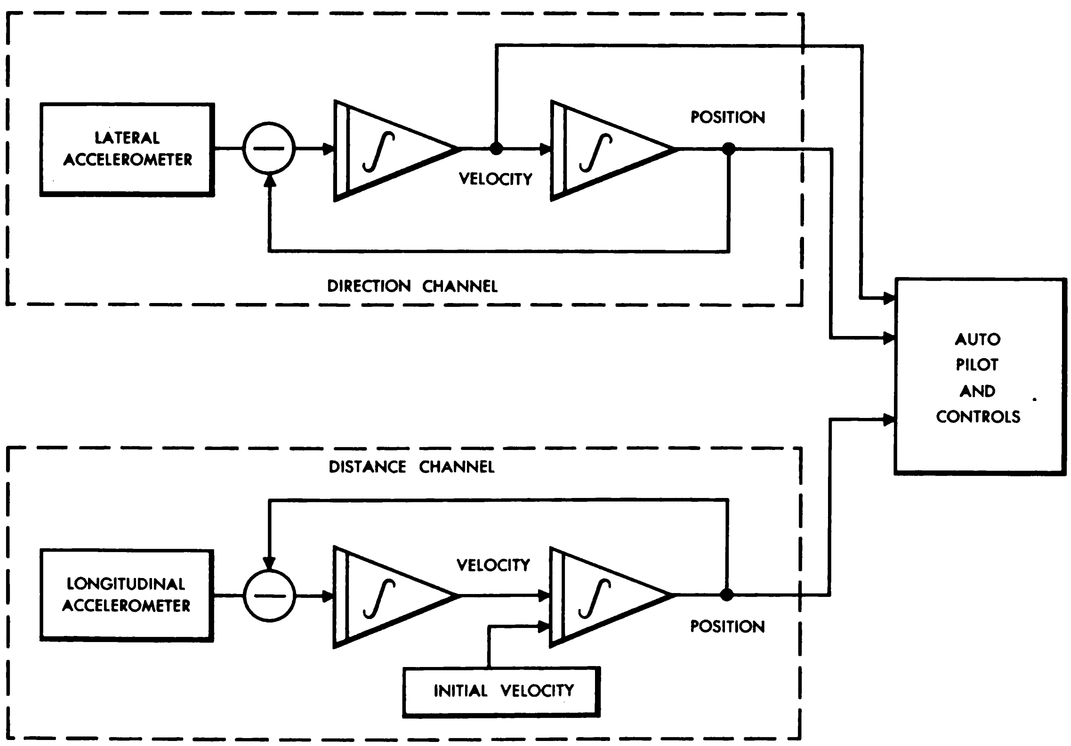
A simple inertial guidance system is shown in block diagram form below.

The system detects errors in the desired flight path by measuring the sideward (lateral) accelerations and the fore and aft (longitudinal) accelerations during missile flight.

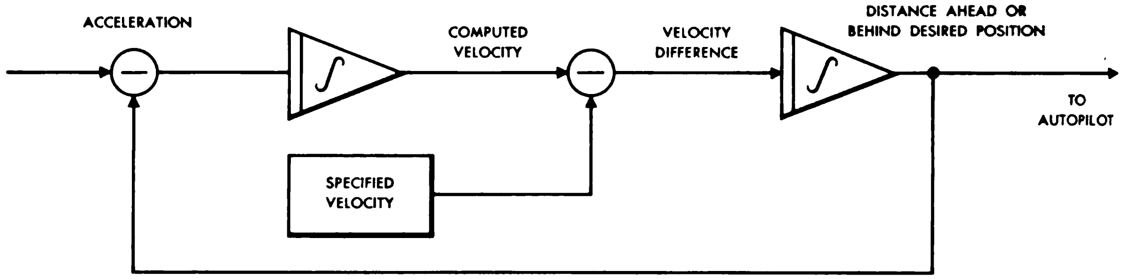
The guidance equipment consists of two main channels: the direction channel and the distance channel. The channels are quite similar in arrangement. Each contains an accelerometer and a method of double integration. The accelerometers detect missile velocity changes without the need of a reference that is exterior to the craft. The acceleration signals are fed to a computer. The computer continuously produces an indication of forward and sideward distance traveled as a result of the acceleration. This is accomplished by integrating acceleration signals to obtain a velocity signal. If this velocity signal is then integrated, the result is missile distance.

DIRECTION CHANNEL. The direction-channel output should be zero at all times if the missile is on course. If the missile drifts off course, the computer output (second integration) is a voltage which compares in amount and sign respectively to the distance and direction the missile is off course. The first integrator of the directional channel has an output voltage moving away from or toward the course. This is a guidance rate signal. Both of these voltages are utilized in the autopilot to determine the amount and direction of control to be applied in a transverse direction to the course. These signals are necessary because the autopilot gyro signal detects heading errors but not off-course conditions.

Because some accelerometer outputs may contain a voltage component due to gravity as well as missile acceleration, some method of separating the two may be needed. The portion of the signal due to gravity is a function of the missile position. If the component of the signal due to this position is subtracted, pure acceleration remains. Only the acceleration signal is to be used in the integrators to solve the distance traveled or position of the missile. With position data already present in the output, it is converted to signals of proper



Inertial guidance system



Computer using specified velocity

scale factor and subtracted from the accelerometer signal. This explains the feedback paths in both channels of the previous figure. The horizontal acceleration of the missile is the remaining signal entering the first integrator.

DISTANCE CHANNEL. The distance channel is identical to the direction channel in operation. The first integrator gives an output whose magnitude indicates the velocity of the missile. The second integrator has a voltage output proportional to the distance the missile has traveled. If the system does not begin operation until the missile has attained a certain velocity, this velocity must be accounted for in the computation of distance. A separate signal representing initial velocity is fed to the input of the second integrator. The output of the first integrator combines with this initial velocity signal to indicate the changes in velocity.

The distance the missile has traveled must be compared to the known distance from launch point to target. A voltage equal to the entire distance the missile is to travel is set up as an initial condition at the time of launching. The voltage is combined with opposite sign to the output of the second integrator. Then, the output of the distance channel approaches zero volts as the flight progresses, and at zero volts the target point has been reached.

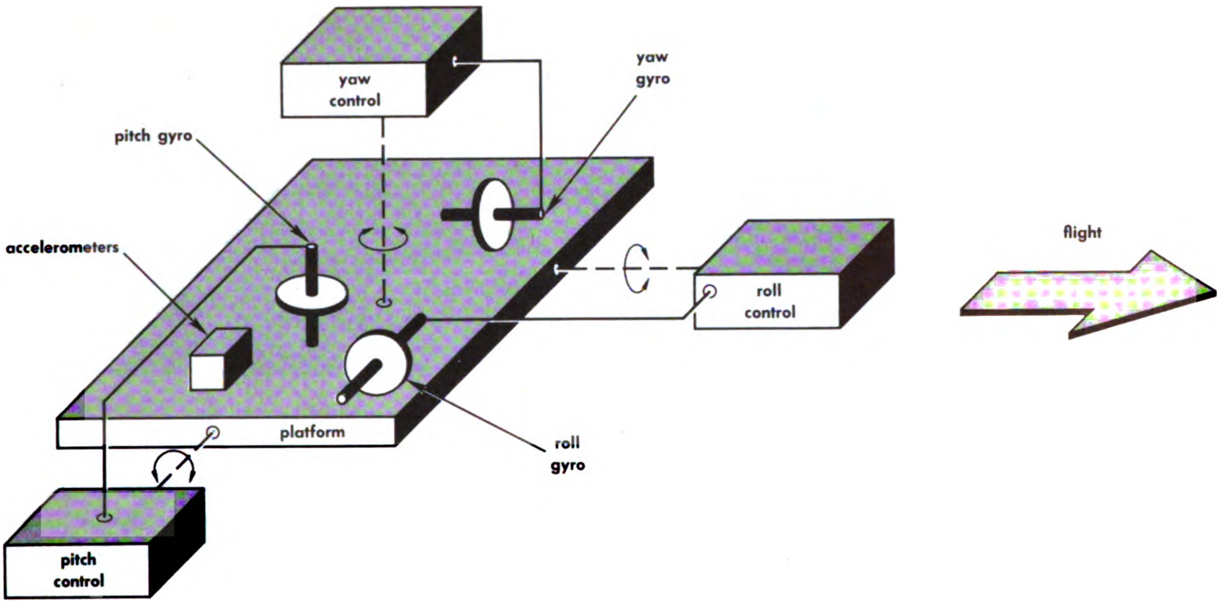
A flight of many thousands of miles would have to produce an extremely large integrator output in order to be an accurate indication of distance traveled. The initial condition of launch-to-target distance would be equally as large. A means of keeping the quantities within practical limits is to continuously specify (program) the distance covered throughout the flight. This desired distance would then be compared to the measured distance (represented by the computer output) in such a manner that both are carried as small quantities. The difference between the specified distance and the actual distance would be a comparatively small quantity. This specifying of distance would be done by a programming device such as an on-board recorder. In such a case, the recorder would determine guidance termination, provided the integrator output is zero.

Another means of keeping signal amplitudes within accurate and reasonable values is to specify the velocity and have that as a reference condition on the first integrator output, as illustrated in the above diagram. Then, any error above or below the specified velocity is fed to the second integrator. This results in an output proportional to the missile error, either ahead of or behind the desired position on course.

Gyro-Stabilized Inertial System

The inertial system described above would be sufficient if the missile were flying straight and level at all times. Aside from errors introduced by the equipment, any variation in pitch and roll would give an erroneous output from the accelerometers. The accelerometers, then, require a stabilized platform so that they will remain parallel to the earth and only detect true errors in acceleration.

In the *gyro-stabilized inertial system* shown at the top of the next page, the accelerometers are placed on a platform that incorporates gyroscopes. They are placed so as



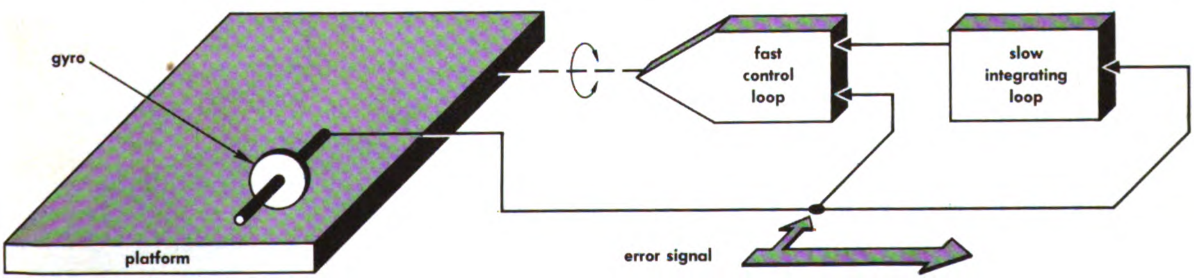
Basic gyro-stabilized inertial system

to detect errors in the pitch, roll, and yaw axes. Any off-level condition is apparent in the output of the gyros. This error signal is then amplified and applied to a servo-mechanism which corrects the platform position.

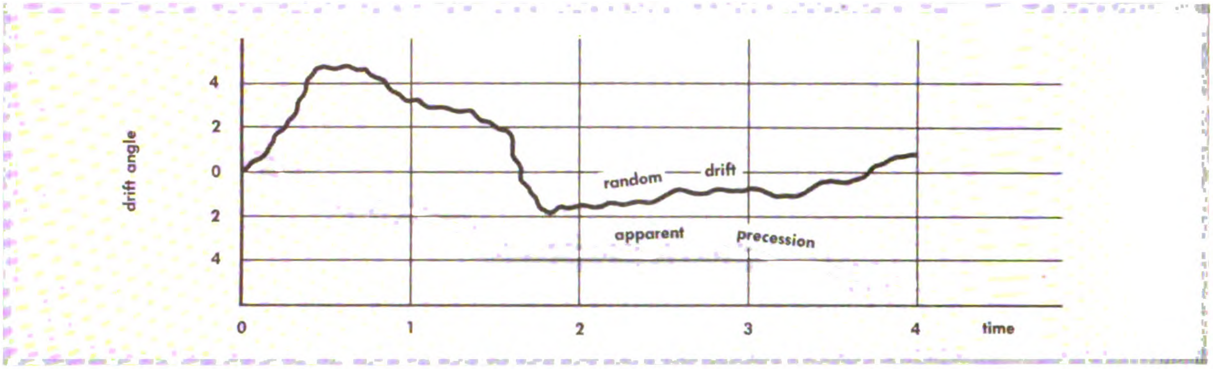
Gyro error. If the gyros had perfect frictionless bearings so that there would be no precessional force to create drift, the above circuitry would be sufficient. Until the art of gyroscope manufacture advances sufficiently, some means of gyro *drift compensation* is necessary. The following drawing shows that compensating circuitry would be required by such a method. This compensation can be added in the form of an integrating (or slow loop) in the servo system. The fast loop and slow loop both have the gyro error output as a signal. The fast loop responds rapidly to correct any deviations in the platform. It operates in a manner similar to the missile stability control system.

The slow loop sums only the gyro drift output error signals during the complete flight, because it cannot respond to rapidly varying errors. Note in the chart on next page that during a normal flight the random drift from a straight-and-level condition is to one side and then to the other. For this reason, a sum of random drift over the whole flight produces a small total error.

The apparent precession as shown in the chart is constantly in one direction. This precession would then continuously give an error output of the same polarity (or phase). Therefore, a summing over a period of time of random drift, plus apparent precession,



Gyro-drift compensation



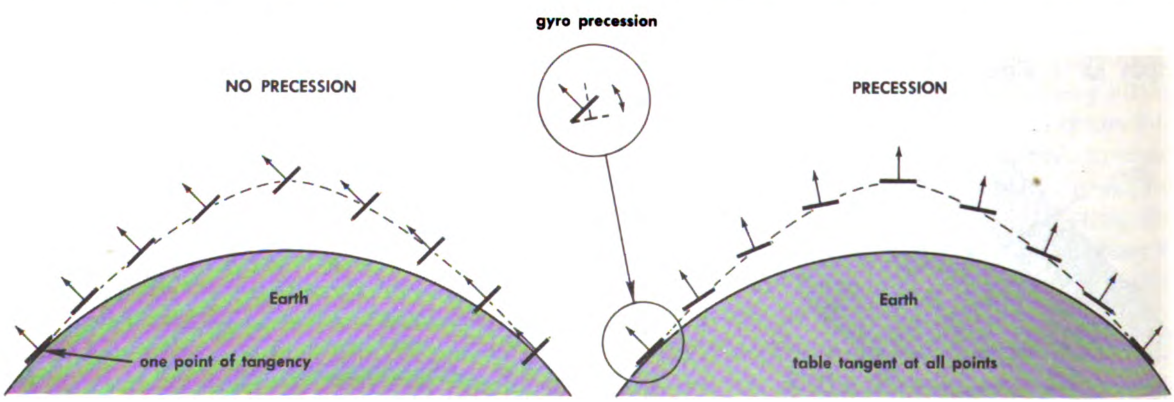
Gyro errors during normal flight

gives an error signal output which is predominantly in one direction. The gyro drift effect is minimized by applying the voltage of the slow loop as feedback to the input of the fast (or control) loop. The slow-loop output voltage is of such polarity (or phase) and magnitude as to cancel the component of gyro output signal due to gyro drift.

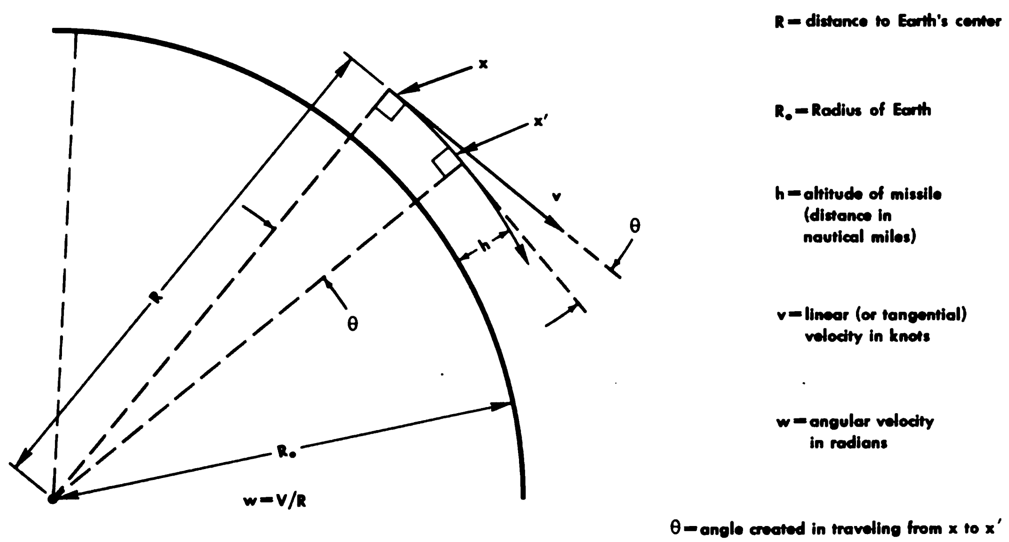
Leveling of a moving platform. Now that the system has provisions for keeping the platform level and preventing drift, a new problem becomes apparent. A normal flight trajectory is an ellipsoid around the earth. A gyro's characteristic of being fixed in space would mean that the platform could only be tangent and upright to the earth's surface at a single place as pictured in the illustration at the bottom of this page. In order to keep the platform tangent to the earth as the missile containing it moves along the trajectory, the forward end of the platform must be depressed at a rate proportional the velocity of the missile around the earth. This keeps the platform level about the pitch axis in respect to the earth's surface, as shown in the illustration to the right.

Usually, gravity is used as a reference for slaving a gyro. But in an inertial system, the platform is kept level by dividing in the computer the measured missile velocity by its distance from the center of the earth; its distance is altitude plus the earth's radius. This results in a factor called angular velocity of the missile in radians. Study the figure above right for the geometric relationship of this velocity. If the pitch angle of the platform is changed at this same angular velocity, the platform will maintain tangency to the earth about the pitch axis.

The platform pitch angle is varied through the addition of circuitry to the missile computer. The pitch gyro is precessed according to the angular velocity of the missile about the



Identical gyros showing single point of tangency with earth



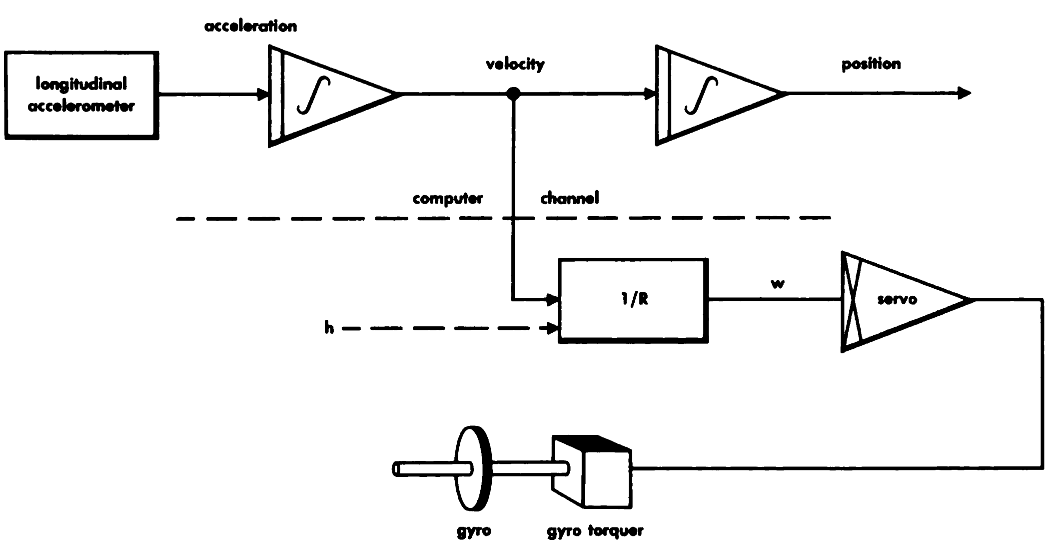
Application of angular velocity to platform leveling

earth. This tilts the platform so that it maintains its tangency with the earth.

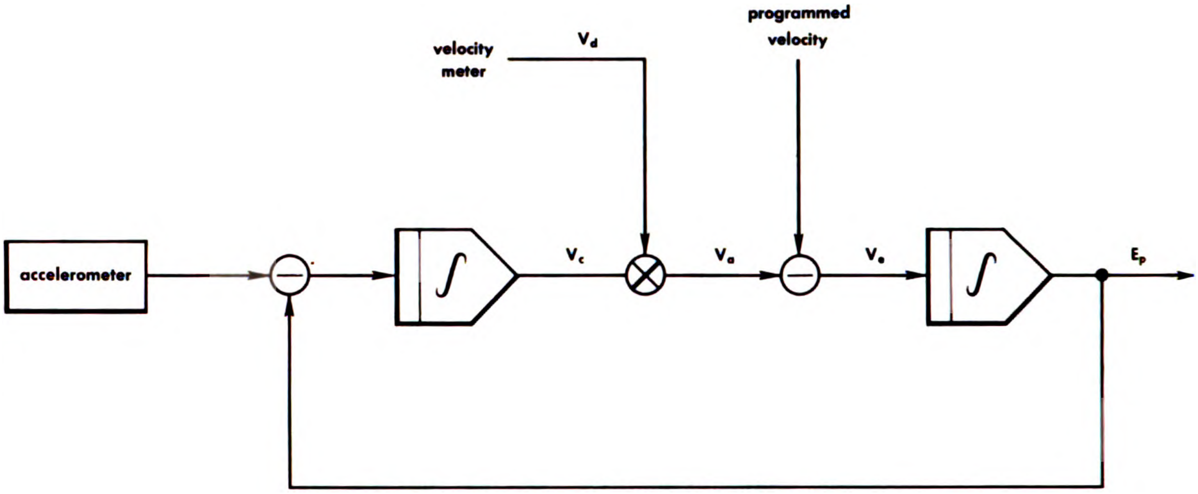
In the figure below, the output of the first integrator, which is proportional to the missile velocity in knots, is divided by the distance (R) to the center of the earth in order to give the missile angular velocity (W) in radians. The angular velocity is used through the gyro torquer to precess the gyro at the same angular rate.

Comparable corrections could be made to motion about the roll axis for side errors. However, the error in tangency would be small because the missile moves such a small distance off course to the side in comparison to the total range of flight. Rather than correct the level of the platform, a proportional bias can be applied to the accelerometer to correct its output signal.

A method separate from the inertial system

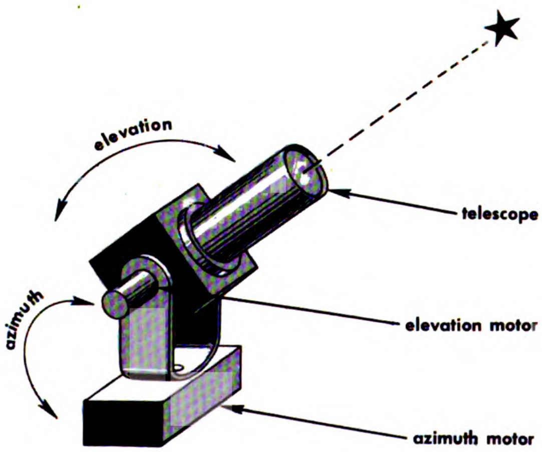


Addition to computer channel for leveling platform by angular velocity



Correcting computer by measured velocity

can be used to determine forward and lateral velocity during the first stages of flight. One method uses Doppler radar. The velocity signal is inserted into the inertial-guidance computer. The purpose of exterior information near the start of the flight is to eliminate oscillation of the missile about the desired course. This oscillation is caused by missile distance and velocity errors in existence when the inertial system begins operation. With outside correction, these errors are diminished as the flight progresses. The craft therefore approaches the programmed conditions more readily. The computed velocity (V_c) and the measured velocity (V_d) are compared in a special circuit, as shown above. The output of this circuit gives the corrected velocity (V_a), which is then compared to the programmed velocity. This produces an error velocity which is integrated to produce position error voltage (E_p). The position error feedback to the input acts to damp out any unknown difference in the accelerometer.



Automatic sextant

This completes the coverage of the inertial principle of automatic navigation. The general form of the system was given, although many differences can exist in the actual circuitry. Previous study of accelerometers, gyros, computers, and other elements has shown that they may take many forms. In a particular system, these devices may be mechanical, electromechanical, electronic, or a combination of these.

Another important thing is that a human must manufacture and adjust these items for proper and extremely precise operation. Any negligence on the part of any member of the team that sends a missile on its mission may result in failure of the mission. It is necessary for each individual to do his utmost to maintain accuracy and high standards of work.

To conceive of other guidance systems as being supervised, inertial guidance does not require too great a stretch of the imagination. Many systems do not have continuous corrections or commands being generated. Rather, periodic or incremental commands to the

system are made. This requires the missile to travel on the inertial carry-over of the last command to the next. The inertial guidance principles that have been developed in this discussion are used throughout many missile systems in some degree. This should be kept in mind, since specific reference to these principles is not made in further discussions of other guidance systems.

CELESTIAL NAVIGATION SYSTEMS

A simplified approach to a celestial navigation system assumes an inertial system supervised in a continuous series of fixes. Let's consider one known as *stellar supervised inertial autonavigator* (SSIA) and another known as *automatic celestial navigation* (ACN).

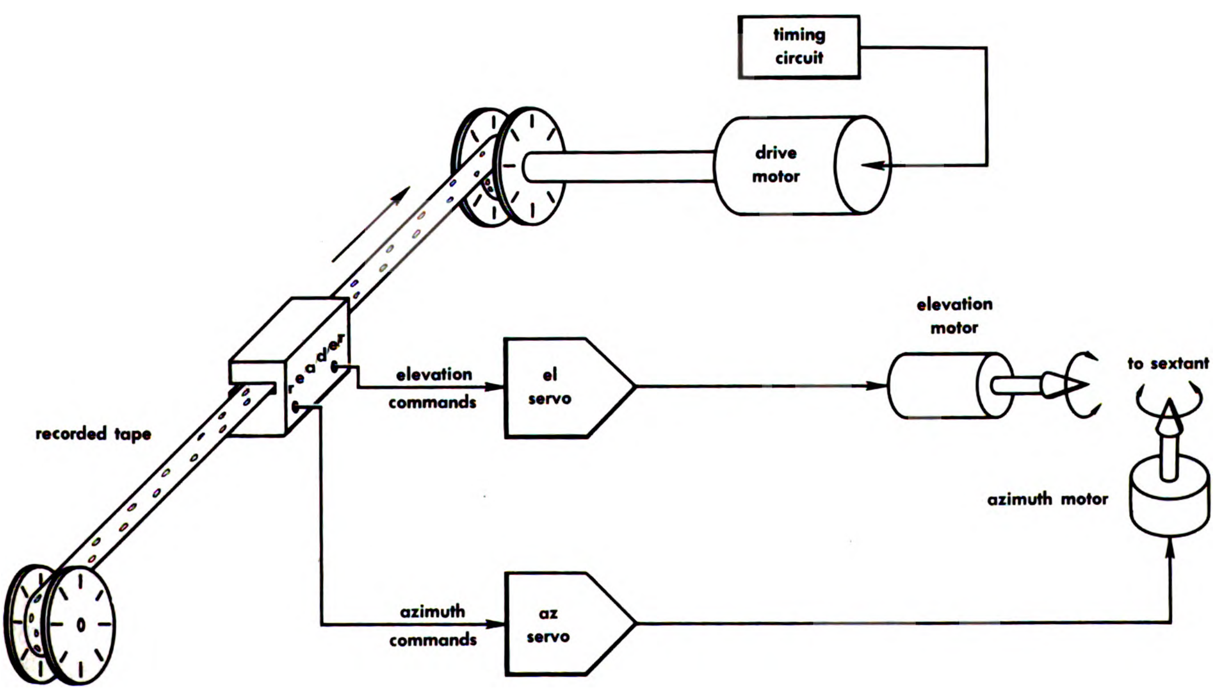
SSIA System of Missile Guidance

In the stellar supervised inertial autonavigator, periodic stellar sights are taken to check on gyro drift. This additional supervision is necessary because not all gyro drift is constant. Gyro drift that is not constant varies as to direction and magnitude. The inability of the slow loop correction to "predict" this random drift leaves an error that tends to increase with the passage of time. For a short flight of 45 minutes or so, random drift would introduce a probable error of about a half mile with modern gyro design. This error naturally increases as the time of flight is increased.

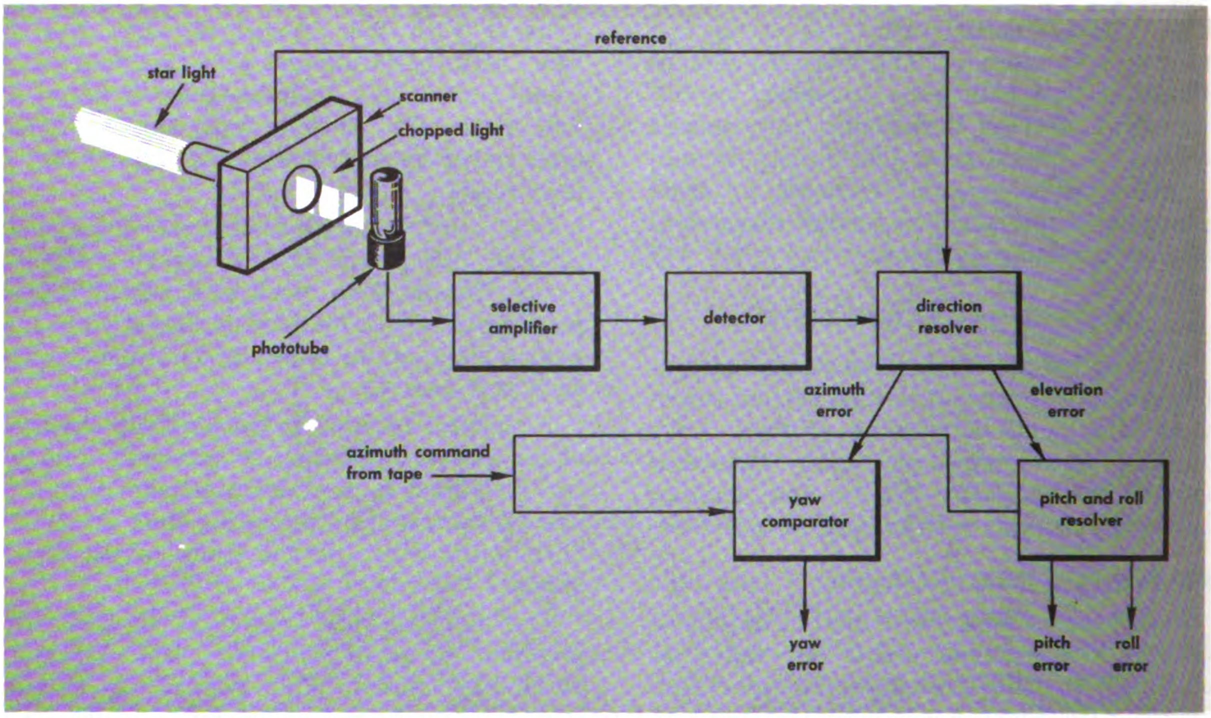
One possible method of overcoming the random drift error is the use of celestial sights. This is done much as a human navigator would check his dead-reckoned position, with a sight on an object of known position such as a star.

A physical addition to the platform consists of an *automatic sextant*, so mounted that it can be varied in the elevation and azimuth axes. A combination mounting that gives the variation is shown in the figure to the right above.

The azimuth and elevation motors are precision torquers connected to a *sextant positioning system*. The system, as shown below, receives its signal from a tape playback which is prerecorded with the necessary position and rate data for the complete flight.



Sextant-positioning system



Stellar error-detection circuit

Telescope azimuth and elevation information must be read from the tape at the proper time. This timing of the tape playback is an important function because a star is at a particular angle with respect to a certain spot on earth only at a particular instant of time.

The telescope is aimed at the star by the information read from the tape and is then programmed through the tape information to follow the star.

A scanning system detects whether the star is centered in the telescope field. The error signal thus developed can be detected and processed to give an indication of the sextant error.

The outputs of the *stellar error detection circuit*, shown above, are several voltages that are proportional to the missile error in pitch, roll, and yaw. After passing through the scanner, light from the star falls on the phototube causing a voltage output proportional to the light intensity. The phototube output passes through a selective amplifier. The selective amplifier separates the signal from the noise. The desired signal is then detected to obtain the information of the error. In the first resolver the detected signal is resolved into azimuth and elevation error signals by comparison with a reference voltage from the scanner. The direction resolver can be a bridge demodulator circuit or some other type of phase detector.

The errors are again resolved to make them usable in the missile's system of coordinates after the azimuth and elevation error in the sextant has been determined. The second resolvers are controlled from the tape reader. The same signal that sets the sextant sets a resolver for the elevation error output. If the elevation signal were not resolved in this manner, there would be no way of determining whether the error were in the pitch or the roll axis.

If the telescope were elevated and pointed directly forward along the missile heading, any elevation error signal out of it would be attributed to error about the pitch axis. Also, if the telescope were pointed out the side in the lateral direction, any elevation error would

be only a function of roll. Therefore, the resolver is needed to determine whether the elevation error signal is due to pitch or roll, or a combination of the two.

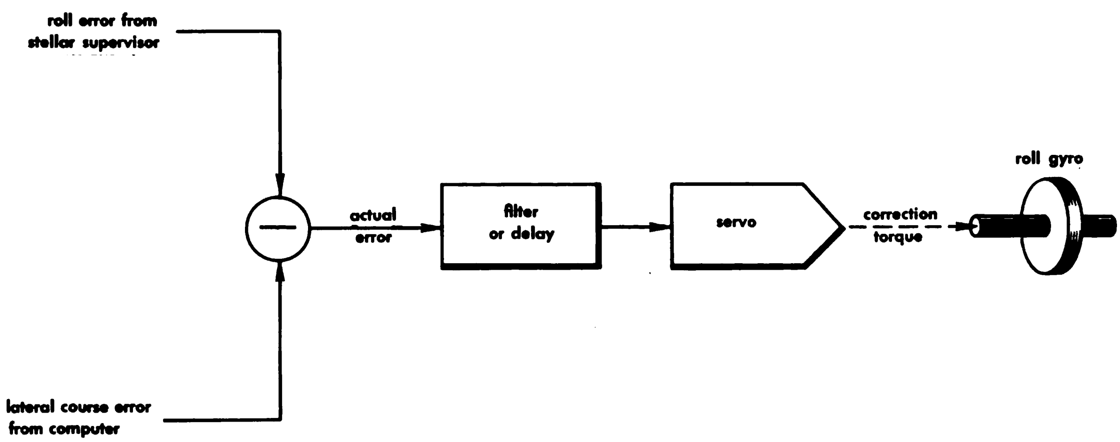
Consideration of the problem shows that there has been no great effort made to preserve the quantity (amplitude) of error signal. The direction and axis of error have been carefully retained but the magnitude has not. Keep in mind that the ideal output of the system is a zero voltage. It would only be an expensive and unnecessary refinement to retain the magnitude; because, in a nulling system of guidance, the error voltage approaches zero and balances out when the missile is on course. Proportional control is not attained. Because of the delaying of the signals through the guidance circuitry and damping by rate function, the missile tends to return to the desired course as fast as possible but without overcontrol oscillations.

An ideal situation in a stellar sighting system would be to check first one star whose line of position (LOP) ran tangent to the course of the missile and then another star whose LOP would be normal to the missile course. In such a situation, the information from the first star would be applied to the computer direction channel. In the second case, the stellar error signal would go to the distance channel in the computer. These signals would then alter the autonavigator operation by correcting the gyros to a new position, compensating for any drift that has occurred. To be assured that the correction torque is accurate, the error signal would have to be averaged over a period of time.

The following figure shows the principle of deriving the difference between the roll error as detected in the stellar supervisor and the course error in the computer. Both errors must be in the same terms; that is, the scale factor of the signals from each circuit must be alike. The feedback loop for such a system is via the new platform position obtained through the star loop and the accelerometer loop back to the computer. If the star is held sufficiently long, the error is damped out to an insignificant amount.

The pitch and yaw gyros are corrected in the same fashion. Some additional consideration for the operation of the yaw circuits is necessary. The azimuth detection of a star is not nearly as accurate as the elevation detection, so in some cases a complex crossmixing of pitch and roll stellar outputs is used to correct the yaw gyro.

With the stellar-supervised inertial autonavigator, the missile can fly for long periods under cloud cover with little effect on accuracy. The difficult part is the development and manufacture of a gyro which can fulfill the requisites of such a system. The techniques needed to produce gyro with zero bearing friction and no unbalanced forces quite understandably tax the ingenuity of the industry. The air-lubricated (air bearing) gyro comes the closest to the ideal gyro for such a system.



Correction circuit for determining true error

The gyros that are easier to produce, and therefore more plentiful, also are more inaccurate. The mass-production gyro is only suitable for use in a continuously (or nearly so) supervised system.

Digressing momentarily, it is not too difficult to conceive that the long-range ballistics missile would have less difficulty with gyro drift. There is a minimum of acceleration forces acting on the gyro to create friction because the ballistic missile is in free fall, thus escaping gravity-induced gyro drift. If the gyro is well balanced, the rotor and gimbals would float in their bearings; thus creating a frictionless condition.

ACN System of Missile Guidance

Based on what has already been covered, the most logical step to overcome the principal problem of a SSIA system would be to use a stellar system having continuous supervision.

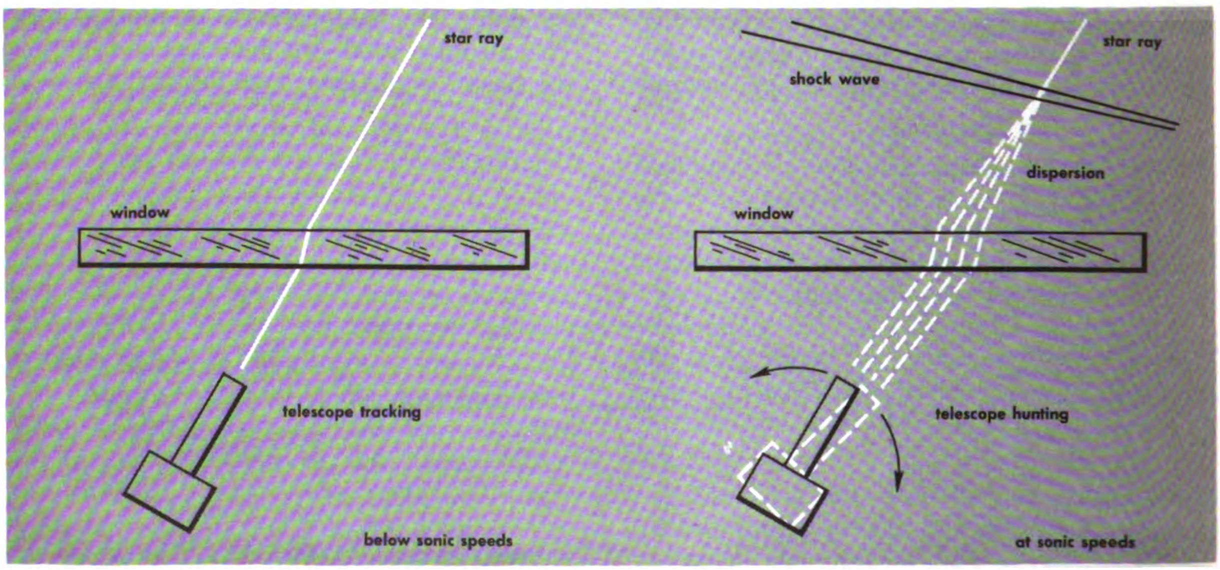
The idea of inertial supervision is usually overlooked in considering *automatic celestial navigation* (ACN) because the system is continually referenced by stellar fixes. But, in any system, the inertial principle still is present in the action of the autopilot between guidance commands.

The platform equipment for ACN requires the addition of one or more automatic sextants. With two sextants operating simultaneously, a series of *fixes* are obtained rather than just LOP. With simultaneous fix on two stars, there is less chance for inaccuracy and less necessity for averaging out errors. A spare or standby sextant might be incorporated into the equipment so that it can be zeroing in on the next star in the navigation sequence without detracting from the other fix.

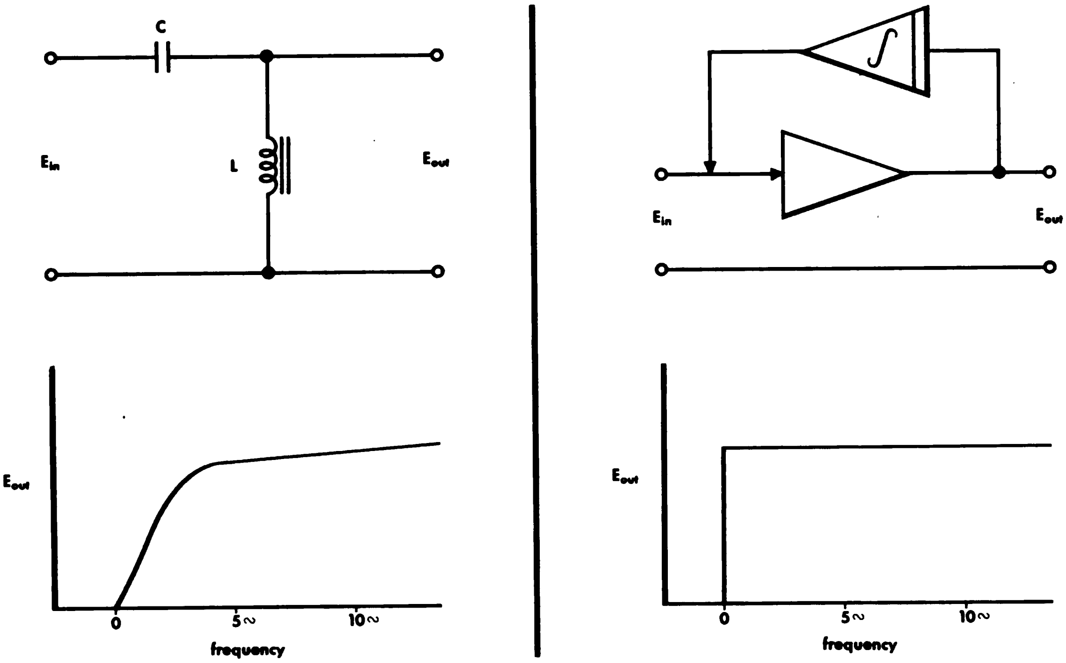
A disadvantage of the multiple telescope system is the need for a window which is large in order to view the celestial sphere. The cost of such a perfect-plane window increases considerably with size. The larger window also finds itself subject to more forces and flections due to high speeds.

Effects of Refracted Light and Bias on Celestial Systems

Refraction of light takes place as it passes through conducting media of varying indices. Since shock waves which occur at or above the sonic speed are, in effect, a conductor of a different index than the ordinary conductor, the dispersion of light rays caused by shock waves may limit the use of celestial or stellar systems to subsonic speeds. However, the subsonic speed limit need only be imposed during a stellar check. Note the illustration below.



Effect of high speeds on stellar systems



Characteristics of two types of high-pass filters

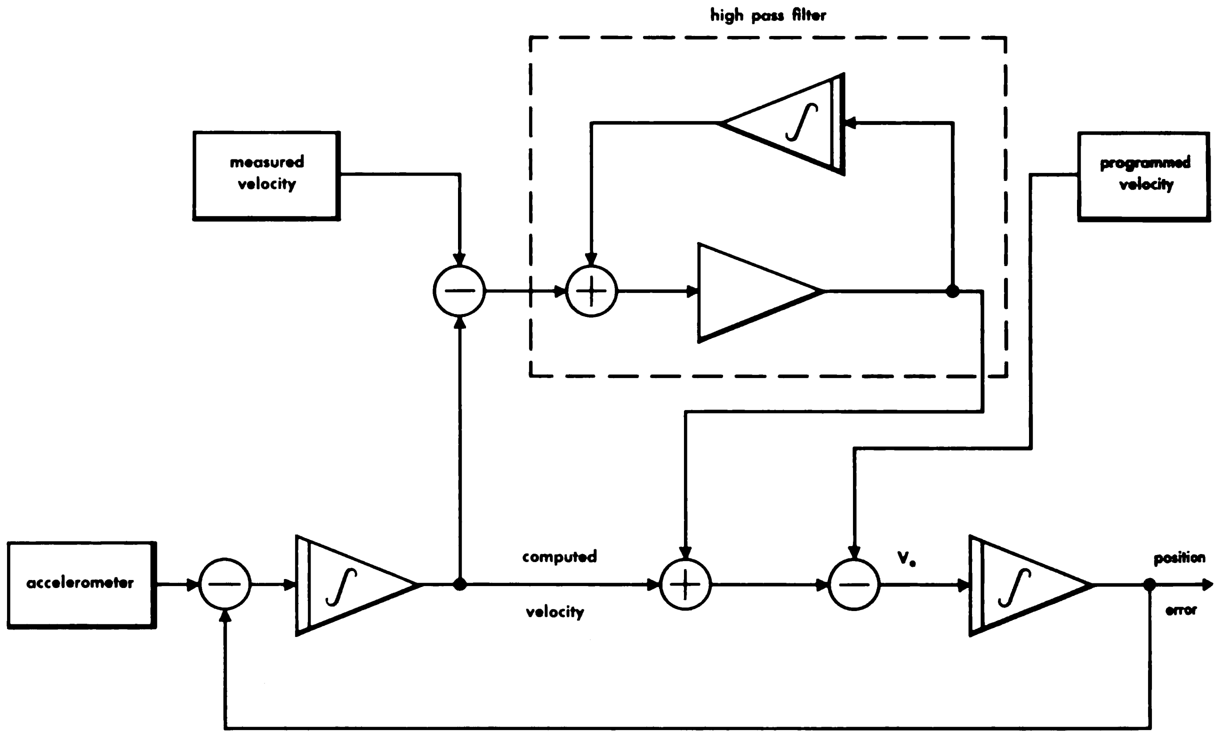
In practical applications, some noise or bias exists on the output of the velocity-measuring component. The noise that exists as short peaks of energy is effectively filtered out by the time constants (delays) of the circuitry. However, steady noises (errors) are not filtered by such action. If some steady error exists on the signal, indicating the measured velocity, then the whole computer output would be in error. The addition of a circuit is required that does not let the bias error through but admits other signals. Such a circuit has to have the characteristics of a high-pass filter which uniformly passes AC current of every frequency. An impedance filter designed to accomplish such a thing is prohibitive to construct. However, a unique arrangement of amplifier elements gives the desired result, as shown above.

The desired high-pass filter characteristic can be obtained using a DC amplifier with integrator feedback. A constant positive voltage (E_i) at the input tends to give a like positive voltage at the output (E_o). The amplifier output becomes the input to the integrator. The output of the integrator with this constant positive input is a negative going voltage. The negative going voltage is added to the input of the unity gain amplifier, where it cancels the input, driving it to zero.

The integrator responds slowly to an input signal. It takes about ten minutes for the integrator output to build up sufficiently to cancel a steady amplifier input. This means all voltages that varied faster than this would overcome the feedback and pass through the circuit before feedback takes effect. These voltages would continue to pass through the circuit as long as they varied at this rate.

In the following figure, a DC filter is applied to an inertial computing channel. Note that the computed velocity which is summed with the measured velocity is varying at the same rate as the computed velocity. In such a case, the filter input would be DC.

Since the varying difference *between* the two velocities is the required factor, the addition of the computed velocity is necessary and creates no error in the connection shown. By damping with externally measured velocity, the summing of errors in an inertial system is eliminated.



Application of error damping

LONG-RANGE HYPERBOLIC GUIDANCE

The utility of hyperbolic navigation exists only for the distance at which the direct or ground wave can be distinguished. Long-range navigation by the Loran principle gives optimum operation on the low frequencies, in the 100-kc to 200-kc region.

Long-range missile guidance equipment, making use of the Loran principle, must utilize an automatic receiver. There is no need for the regular indicator unit, however, since the outputs must be in the form of shaft rotations or voltages usable in the control systems.

Pairs of transmitting stations (master and slave) generate the pulses of RF energy separated by a delay interval. The delay exists after the master pulse. The delay exists while the pulse signal travels to the slave site and is used to key the slave transmitter. This delay is used to insure that the transmitted pulse of the master station is always received first at a receiver located in the area serviced by the system. Pulses of a particular time difference set up a LOP that is hyperbolic in nature. In the functioning of a Loran system, the pulse time differences are first measured for the coarse indication. Thus, the beginning RF cycles of the two transmitted pulses (master and slave) are compared for the fine range indication. In a manual-type Loran system, this matching is done by the operator using the presentation on a cathode-ray tube equipped with a precision sweep. To accomplish this automatically requires a more complicated receiver.

Some method of minimizing noise should be added to assure accurate operation under as wide a range of conditions as possible. Locally generated signals which are equal in time, frequency, and phase to the transmitted pulses are used. These local signals, which are noise-free, are compared to the transmitted slave signals, giving a better and more accurate comparison.

Automatic Long-Range Loran

An automatic phase-matching type of Loran system breaks up doubtful cycles of informa-

tion, and it separates them by transmitting alternately on two different radio frequencies. The pulse envelope timing and the phase of both radio frequencies are controlled by the transmitting timer. By proper adjustment of phase and envelope timing, the envelope match and the RF match on both frequencies can be satisfied simultaneously in the receivers.

Shown on the following page is a simplified block diagram of the receiver for the automatic phase matching system. Servo loop 1 operates to bring the 25-cycle pulse generator into *time coincidence* with the 25-cycle modulation envelopes of the received master signals of 180-kc and 200-kc carriers. Servo loop 2 operates the variable delay system to bring its delayed output pulses for slave gating into coincidence with the received slave signals. Thus, the cycle matching on the master signal involves comparing the noisy master signal with the noise-free local reference signal. The same operation is carried out with the slave signal. This permits operation under worse signal-to-noise conditions than in a manual, visual, long-range Loran system. The amount of delay is indicated on the Veeder Root counter, and this is the envelope time difference reading between master and slave pulses, furnishing a coarse indication of the LOP's.

Servo loop 3 brings the 180-kc CW reference wave into phase match with the received master 180-kc signal by shifting the phase of the 100-kc oscillator output. The 180-kc CW reference and the received 180-kc slave pulses are fed into the servo loop. The phase shifter is driven to bring these waves into phase match. Thus, the rotation of the phase-shifter shaft of loop 4 represents the phase difference between 180-kc master and slave pulse signals.

In a similar fashion, servo loop 5 sets up a 200-kc CW reference wave which is in correct match with the received master 200-kc RF pulses. As with servo loop 4, servo loop 6 drives its phase shifter to bring the two inputs of the phase-difference detection system into matching phase. Again, the shaft rotation of the phase shifter of servo loop 6 represents the phase difference between the received master and slave 200-kc signals.

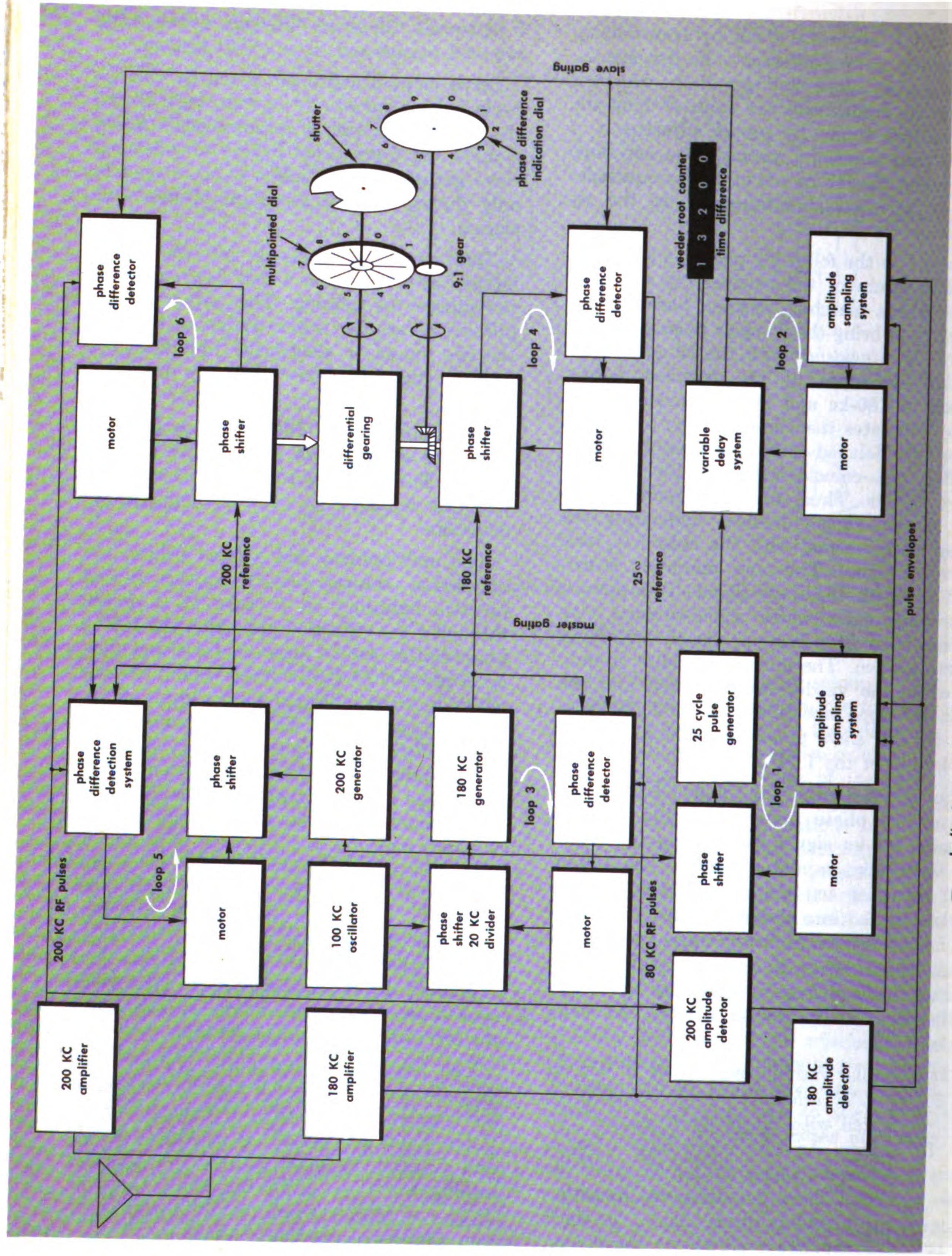
Servo loops 3 and 5 are gated by the 25-cycle pulse generator to operate only during the time of reception of master signals. Servo loops 4 and 6 are gated by the time-delayed slave gating 25-cycle pulses so as to operate only during time of reception of the slave signals.

The mechanical outputs of the phase shifters of servo loops 4 and 6 are fed into the differential gearing. The output of the differential gearing is the difference between the two inputs or, electrically, the difference between the two phase-difference measurements.

The shaft of the differential output has a circular shutter with a 40° sector opening cutout. This opening represents the 20° of tolerance required for cycle-difference measurements.

The multipointed dial located behind the differential shutter has 9 equally spaced pointers indicated on a scale. It is driven through a 9:1 reduction gear from the 180-kc phase-shifter shaft. Consider a fractional nine-cycle interval of the received time difference which is to be indicated by the cycle difference and the phase difference indicators. The phase difference or the fractional cycle difference is indicated on the phase-difference indicator. The number of whole cycles is on the multipointer dial. The shutter, and its 40° opening, indicates the presence of errors in the phase-difference measurements by the position of the arrows of the multipointer dial with respect to the shutter opening. No error due to low signal-to-noise ratios exists if the arrows are present. If no arrow is present, errors sufficient to prevent correct cycle-difference measurements are present in the phase-difference measurements.

The positions of the phase shifters that indicate the intervals can now be utilized in a computer. The information from this computer could be utilized to obtain distance and direction to target, to plot true position automatically, to control the automatic pilot, or to give whatever information of this order that may be required in the system.



Indicator-receiver for automatic phase-matching system

The base line for a system accurate within a few hundred feet at 1500 miles range would need to be 600 to 800 miles long. The size antenna required for the receiver in a missile would also impose some problems. An adequate receiving antenna would need to be 30 to 50 feet long, or the equivalent. Using sections of the missile airframe to fulfill this purpose is possible.

FM Loran

An FM Loran system is similar in practice to the one just discussed, but it uses a unique approach to eliminate ambiguities. In this system, the outputs of three transmitters are *frequency modulated by a sine wave.*

These three transmitters, with different low-frequency carriers, are frequency modulated by the same AF signal so as to obtain identical modulation in frequency and phase for the three transmitters. The time for one cycle of the modulating frequency must be sufficient to allow the RF signal to get to maximum range of the system. The receiver automatically tunes out the sky waves. Because of the longer path traveled by the sky waves, the signal produced by them has a different frequency than the ground wave at any instant of time.

Comparison between a pair of transmitted signals is accomplished by measuring the relative delay required to produce a phase match in their modulating functions. The phase match is indicated by a constant difference-frequency output of maximum amplitude from the mixer. The phase of this difference-frequency is also an indication of the RF phase difference of the two signals being compared. Phase changes in this difference signal can be detected by comparison with an equal frequency signal locally generated. In this manner, a method of fine control is provided by RF cycle matching. Time difference is measured by the phase shifts required to maintain the phase conditions for signal pairs.

A reference oscillator system, using only two base stations but requiring a precision reference oscillator in the missile, integrates the time difference to measure range absolutely. It thus establishes circular lines of position about the two base stations.

The term *precision reference oscillator* is a slight understatement. The requirement for this airborne oscillator is such that it starts its cycles at the regular period intervals throughout its operating life. Practical oscillators have a slight phase angle with changes in load, power supplies, temperature, or physical conditions. Any slight phase shift in this system would create error that would be difficult to detect.

In the final analysis, only one of the foregoing long-range hyperbolic automatic navigation systems will be accepted. Development of these systems has been going on for several years; and, theoretically, either system can accomplish the same purpose. The system which wins out will be the one for which the most accurate and dependable equipment is developed.

TERRESTRIAL REFERENCES FOR LONG-RANGE GUIDANCE

Various picture and mapmatching guidance systems have been suggested and devised. The main idea of an electronic unit of this sort is the comparison of a photo or map contained in the equipment with an image of the area the missile is flying over.

If a negative and positive transparency of a scene were to be matched and held up to the light, the combination would appear opaque. Now if either transparency were moved somewhat in respect to the other, light would show where images were out of register. If one transparency were contained in a frame activated by a servo, it would be possible to devise a detecting control mechanism that would automatically rematch the images. Instead of a positive transparency, the projected image from a lens or a radarscope could be used to show the actual picture of the area being traversed.

At this point, it would be best to rule out daylight systems, that is, the use of photographs of the actual course or target area. A daylight system would put restrictions on operations due to visibility conditions and would limit operations to a time when interception by target defenses would be simpler. A radar mapmatching system would have no limitations as to the effects of weather

or darkness; however, it has some other drawbacks which are covered later.

Magnetic guidance is another type of midcourse terrestrial guidance used in missiles. This type is discussed after the mapmatching systems.

Radar Mapmatching Systems

The block diagram on the following page illustrates a workable guidance system which utilizes radar mapmatching. Although more adaptable as a homing system, radar mapmatching can be used for long-range guidance

The radar map comparison is made by projecting the PPI image, by means of a rotating offset lens, through a negative transparency of the same region and onto a photo multiplier tube. When the PPI image coincides with the map image, the light transmitted is minimum. The map is printed on a strip of film and the lens rotated, causing the PPI image to be moved in a small circular pattern on the film. The effect produced in the photo multiplier tube output is similar to that obtained by an offset radar scan. When the output is properly commutated, left-right and fore-aft information is obtained.

The commutated pulses from the output of the photo multiplier tube are applied to DC discriminators and integrators; then, following through the block diagram, this information is supplied to two loops, lateral and longitudinal. The left-right information is applied to a servo amplifier which drives the film carriage laterally to maintain the match. The position of this carriage is picked off as an error voltage for the control system. And, as the missile turns on its yaw axis, the film carriage is moved and the error cancelled out.

The fore-aft information is applied to the longitudinal servo loop which pulls the film through the holder at the correct speed to maintain the match. Thus, the film is moved at a rate proportional to the groundspeed of the missile. The film may be keyed to indicate the location of a change of course or to initiate dive action.

Errors in indicated longitudinal position can result from a difference in altitude between reconnaissance and tracking runs, because of slant range distortion and altitude-return delay. The error is greatly decreased if the altitude error is compensated for in the longitudinal and lateral loops. It is known that altitude and azimuth errors produce lateral and longitudinal errors, which are sine and cosine functions of the antenna (vertical) scan angle. These false error signals cause an interaction between the loops. The interaction varies with different loop gains. Altitude and azimuth error signals are not derived conventionally from an altitude or azimuth discriminator. Altitude signals come from the lateral discriminator and are fed into the altitude error detector. The altitude integrator uses this output to furnish the altitude compensating circuit with a signal for expanding or contracting the video presentation. The video presentation is expanded or contracted to make it correspond to the scale of the transparency. The azimuth signal is derived in the same sort of functional circuitry, but its output is utilized to orientate the antenna horizontal scan on the proper bearing.

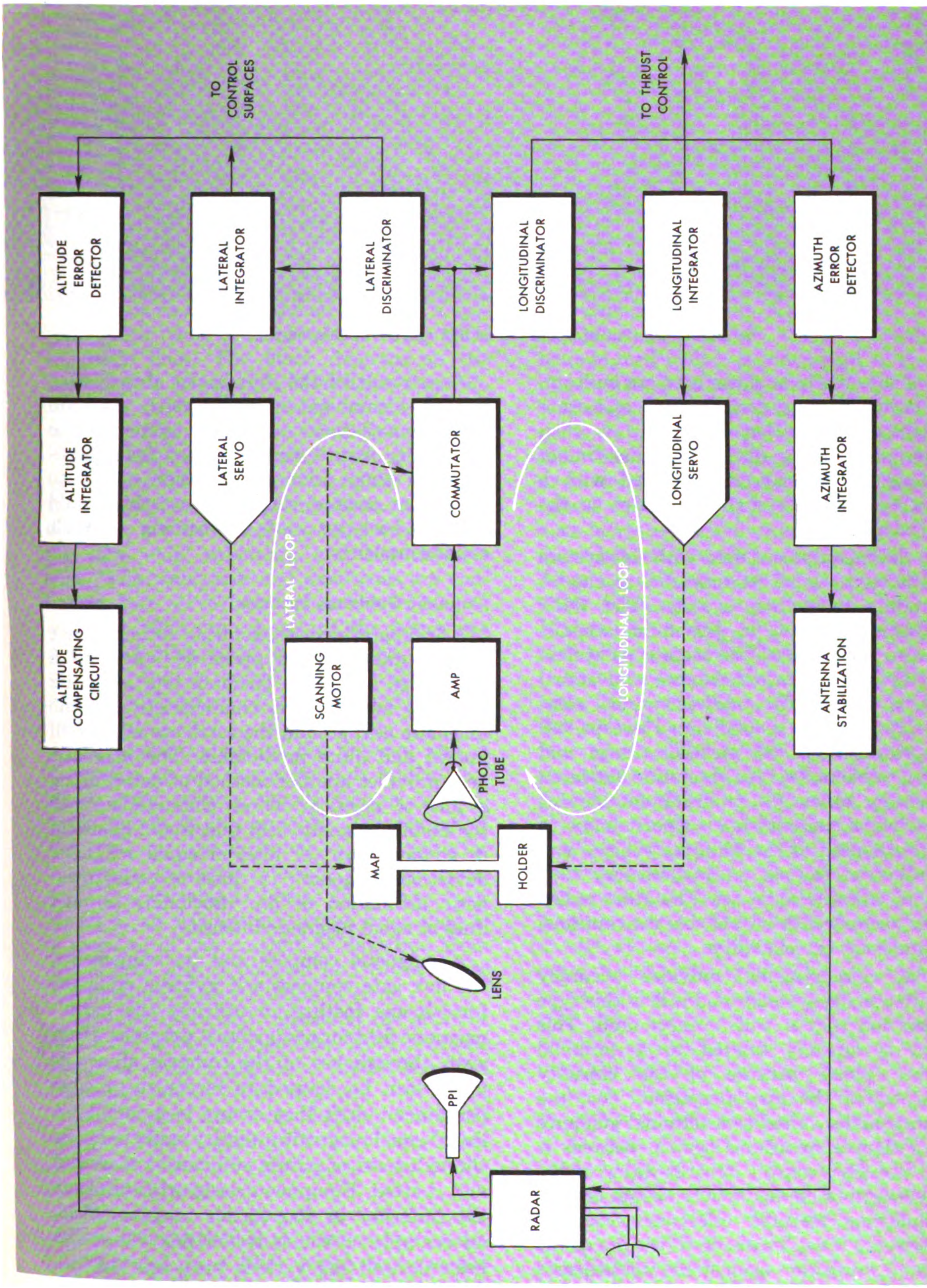
It is necessary to have good angular match (within about 1 degree) before accurate left-right and fore-aft information can be obtained. The original angular match can be obtained by means of a magnetic auxiliary such as a compass. The match is maintained by the azimuth loop of the system.

Two types of holders for the film are possible. A frame-type film holder is the larger and more complicated, mechanically, of the two. It switches separate frames into the system and affords an easier initial lockon. However, the best method for investigating the filmstrip seems to be by scanning it through a mask with a semicircular opening, as shown on page 446.

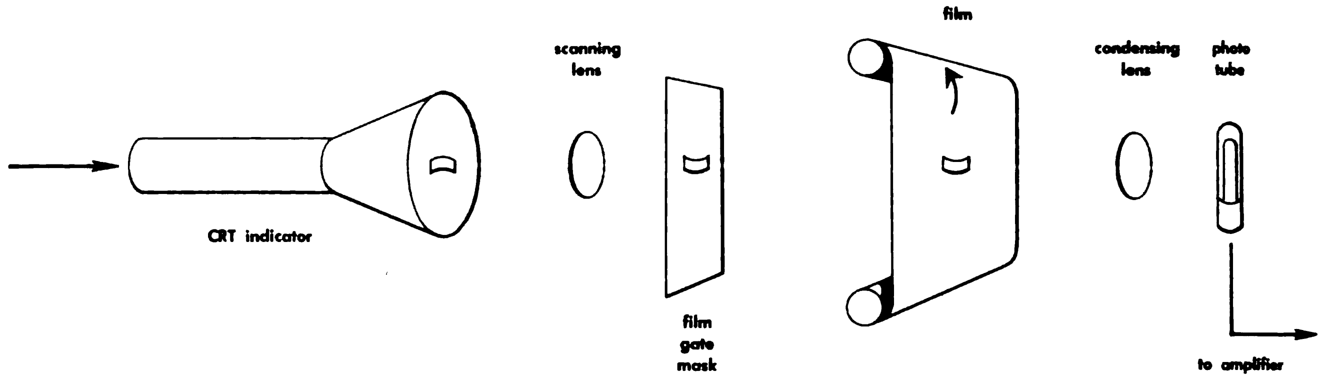
The filmstrip is pulled at a rate corresponding to the groundspeed of the missile. The length of a strip may need to be only about 1/20 of that of the frame-type map.

The radar maps used are the result of two procedures. Of these procedures, actual mapmaking flights over the terrain to be used is the best primary method. However, the actual reconnaissance of a target area, a procedure required by this method, may be difficult; so synthetic maps are often used.

Synthetic *thiokol* base compositions of relief



Radar mapmatching system



Schematic drawing of operation of film holder for strip map

maps of the area to be flown over are built up, using ordinary maps, aerial photos, and other intelligence information. The map is then photographed, using ultrasonic radar-trainer techniques. This map can be used with a success that is only slightly inferior to the actual maps.

The radar equipment used with this system would have the same requirements as one for a bombing navigation system. The CRT indicator would have to give minimum distortion to the picture. The same type system would be used in reconnaissance and actual matching flight. This indicator unit must have: (1) precision ground range sweeps, (2) PPI presentation, (3) precision range and azimuth markers, (4) high-resolution flat-faced cathode-ray tubes, (5) provision for moving the indicator presentation proportionally to the groundspeed of the airborne vehicle, and (6) provisions for use as an accessory to a standard airborne radar set. The other radar-system components would be conventional with emphasis on reliability.

Radar mapmatching provides accuracy equal to homing systems, and it is in this type of operation that it will be most used. This system is *logically* limited to a terminal guidance phase because of the capacity limitations of the map magazine. Too, the system is better suited as a terminal guidance system than a midcourse system because it would not give control on over-water flights of any length or on flights over terrain lacking distinguishing features. Electronic countermeasures are a big drawback to the use of the system; a protection against countermeasures is the highly directional antenna.

As a result of the drawbacks, this system is used in combination with a fairly accurate and nonradiating midcourse system. The terminal guidance would be used for a minimum of time prior to explosion so as to afford the greatest surprise, since this would be the best way of overcoming possible countermeasures.

Magnetic Guidance Systems

A comparatively economical, although not too accurate, system *primarily suited* to a long-range midcourse guidance is a magnetic guidance system. The characteristics of the magnetic field around the earth are fairly well known and quite predictable. The magnetic field provides another method of determining a line of position. Three characteristics of the earth's magnetic field that would be useful in guidance are (1) lines of equal magnetic deviation (isogonic), (2) lines of equal magnetic inclination or dip (isoclinic), (3) and lines of equal magnetic intensity.

A magnetic compass and its refinements, the flux gate compass and gyrosyn compass, furnish one method of utilizing the earth's magnetic field for navigational purposes. However, through use of the compass alone, no means can be devised to recognize or compensate for drift that occurs in a missile. A line of equal magnetic intensity exists uniquely through a set of points and can be measured and charted. Equipment designed to measure the intensity of the earth's magnetic field, an example of which is the flux valve, has been devised for various purposes and can be utilized in missile guidance systems.

With an external magnetic field present, the signal output at a flux valve is the second harmonic of the excitation frequency. Without the external magnetic field, no second harmonic signal would be evident. The strength of any second harmonic output is proportional to the intensity of the external magnetic field. Of course, the foregoing only holds true if the coil axis of the flux valve is properly aligned with the lines of force of the external field.

Now let us consider the structure of a system which could guide a missile using these magnetic lines. The block diagram on the next page is an example of such a system. To measure the total intensity of the earth's magnetic field, three flux valves are required. These elements are aligned along three mutually perpendicular axes and rigidly fastened to one another. These elements, shown in the diagram, are known as the transverse orienting, the axial orienting, and the detector coils. The transverse and axial coils are connected into servo loops in such a way as to null out. The nulling out of the transverse and axial coils with respect to an external magnetic field leaves the detector coil oriented along the direction of the earth's total magnetic field.

The signal output of the detector coil is connected to an electronic unit which produces error voltage signals (shown as 400-cycle signals to the controls). The output of this unit may be adjusted to zero volts for any given magnetic intensity by the adjustment of a variable DC bias. The output error signal is an AC voltage whose amplitude is a measure of the amount of deviation from the desired line of position and whose phase is a measure of the direction of error.

The electronic unit consists of the detector and axial orienting channels. Note that the oscillator supplies signals of the fundamental frequency to all three channels. Automatic volume control and filter elements which are used to insure signals of constant amplitude free from harmonics are incorporated in the stable oscillating unit.

The detector coil output, distorted by the effect of magnetic fields described previously, is connected to a filter which removes the fundamental frequency. The resultant second harmonic signal is connected through an amplifier to a frequency divider, which converts the signal frequency to that required by the control system.

A source of regulated DC, shown being fed into the detector coil, is provided to balance out part of the earth's magnetic field and also to provide the zero adjustment for any field-strength measurements.

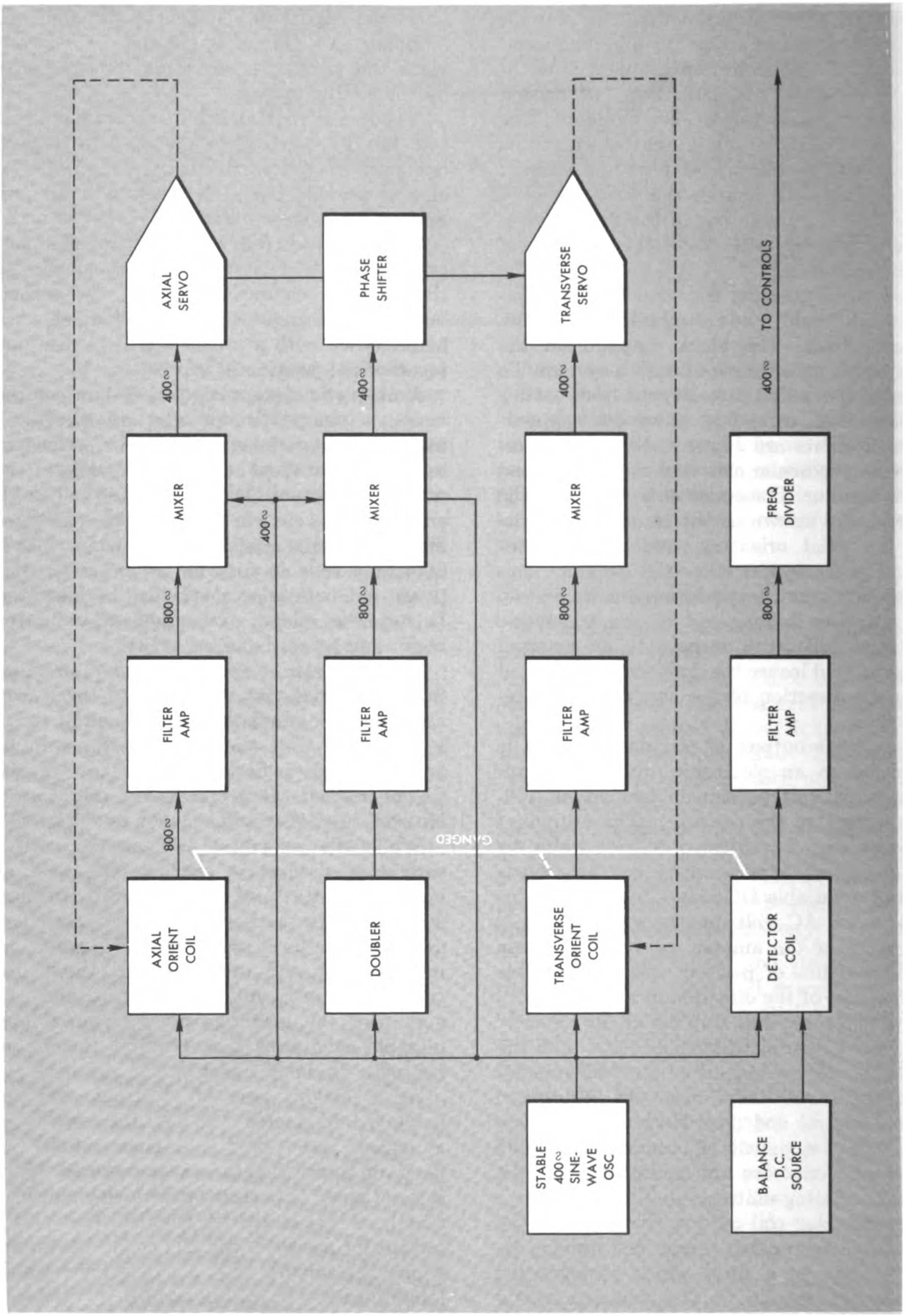
The axial and transverse orienting channels are identical in structure, as you can see, to the detector channel; however, the second harmonic outputs of these channels are heterodyned with a proper signal to feed one phase of the positioning servos.

A magnetic system consists of two components, a magnetometer and an electronic unit. The magnetometer would be so located as to isolate it from stray fields arising in the missile equipment. Tactically, magnetic guidance increases system security. For midcourse flight it permits a silent missile which makes detection difficult and, for all practical purposes, eliminates the possibility of jamming. It has the added advantage of unlimited traffic and target-handling ability.

Accuracy can be good up to about 7 miles from the target but is limited to the course line only. No method of determining range is available. The missile would have to be launched near, or flown to the vicinity of, the line of intensity that crosses the target area. No evasive action could be taken by the missile.

While the magnetic field at the earth's surface is accurately charted, the field at higher altitudes is not so well known. The local magnetic irregularities would have less effect at high altitude, so the earth's field would be more regular and predictable at these altitudes. However, magnetic storms and the area near the magnetic poles create unpredictable conditions that prohibit the use of magnetic guidance under such conditions.

As is always true of electronic systems employed in guided missiles, the long-range guidance system used in a given missile depends on the mission to be accomplished. Thus, it is not possible to state here which long-range system is best suited for missiles. The job intended for a missile and the conditions under which the missile must fly determine which long-range system will be used to guide it.



Magnetic guidance system

Terminal Guidance

We now have come to the final phase of missile guidance — terminal guidance. You are already aware that either a short-range homing system or an inertial system is used for the final phase. Let's first consider homing systems.

HOMING GUIDANCE SYSTEMS

A missile homing guidance system is one which can "see" the target by some means and then institute commands to the control system so that the missile can fly to the target. There are three classes of homing systems: *passive*, *semiactive*, and *active*. If a target illuminates itself, the system used to detect the target is known as a *passive* homing system. If the target is illuminated by some source other than itself or equipment in the missile, the system is known as a *semiactive* homing system. If the target is illuminated by equipment in the missile, the system is called an *active* homing system.

Another way to divide homing systems is by the frequency spectrum to which the system is sensitive (seeks out). A brief description of the various types of seekers follows:

Moving through the spectrum from low to high frequency, we note that *sound* has had some use in seeker systems. Naval torpedoes have been developed as passive sound seekers, but such seekers have certain drawbacks. The sound-seeking missile is limited in range and utility because it must be shielded or built so that its own motor noises and sound from the launching point will not affect the seeker head.

Electromagnetic radiations are the most popular of the media of the homing systems. *Radio* has use in the *passive* homing system. The seeker acts as an automatic direction finder on a frequency being transmitted from the target area, and it homes on that frequency. There is no weather or visibility restrictions, but it is unlikely that there would be a radio transmitter conveniently operating in a target area. Radio jamming can do a thorough job of "confusing" such a unit.

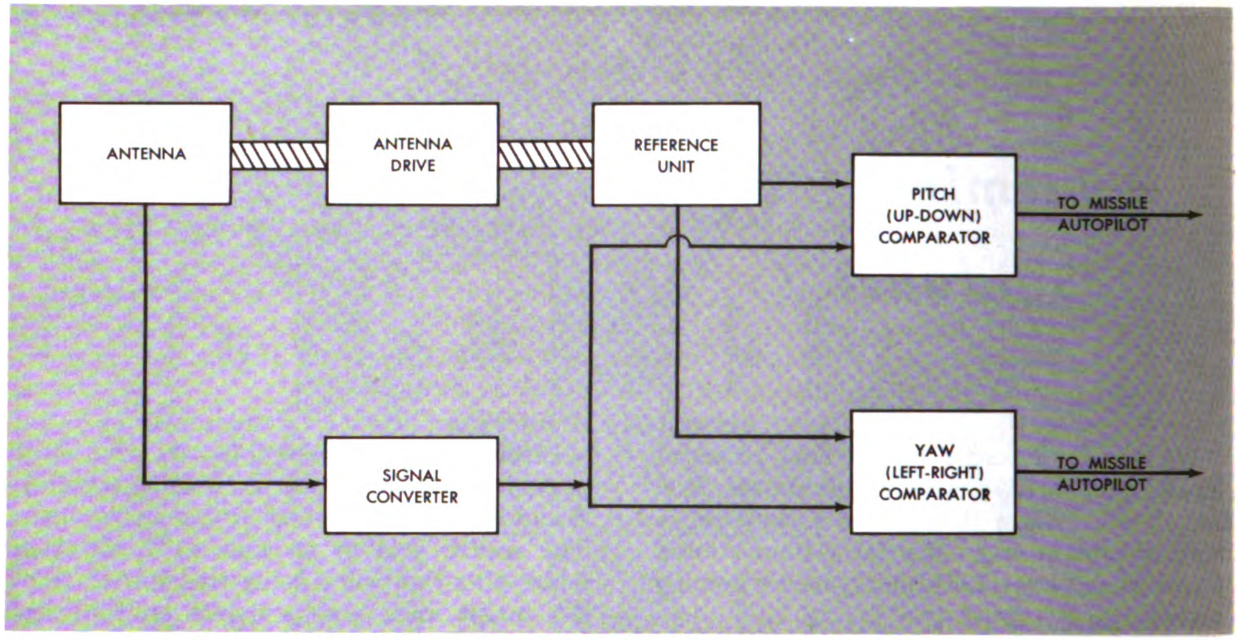
Radar can be used in any of the three classes of systems but is best suited for *semiactive* and *active*. At the present this use of electromagnetic radiation in a target seeker is foremost in effectiveness. Radar is little restricted by weather or visibility; but it is susceptible to enemy jamming.

Heat is best used with a *passive* type seeker. It is difficult to jam or decoy heat-seeking systems when used against aerial targets because the heat developed by engines and rockets of the aerial targets is difficult to shield. Providing a sufficiently sensitive sensor can be developed, this system has great promise.

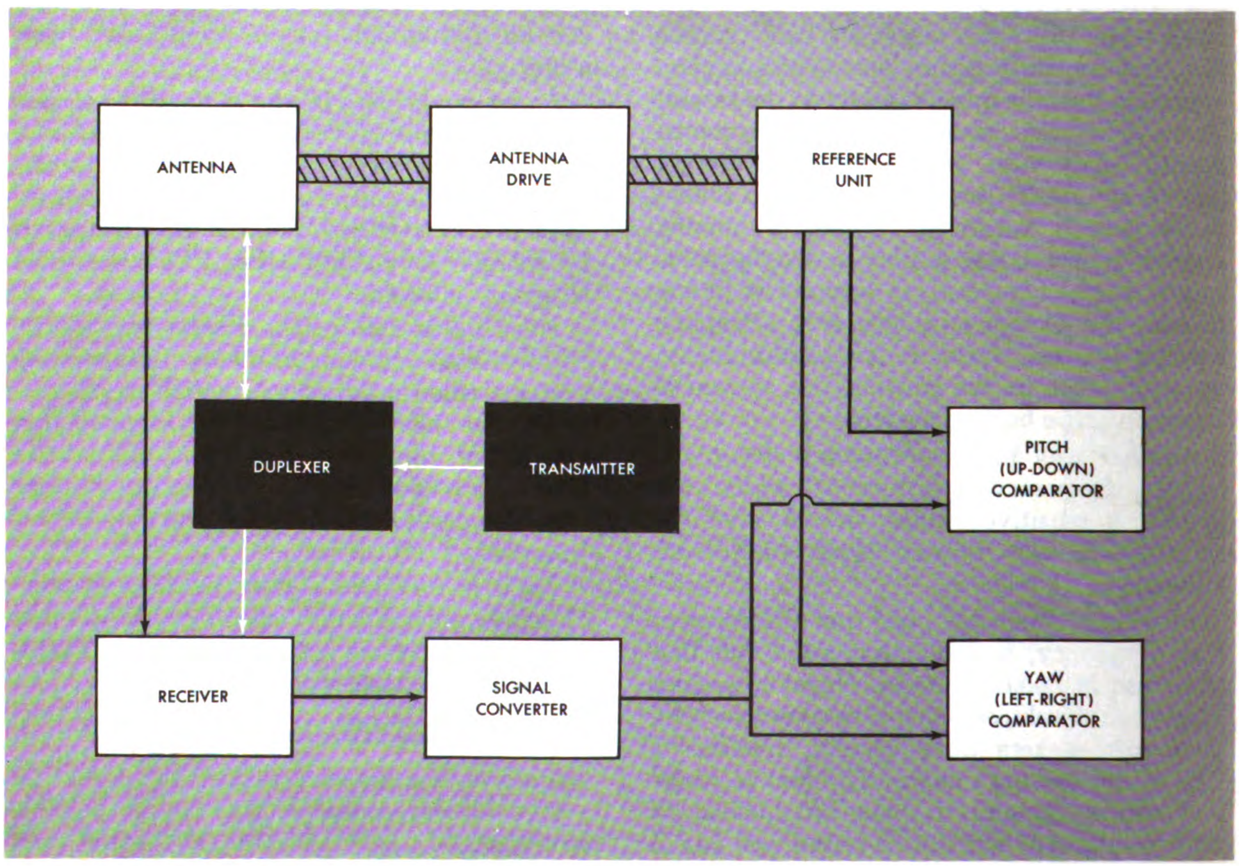
Light is useful in a *passive* seeker system. However, its use is restricted by both weather and visibility. Such a system is quite susceptible to countermeasure techniques.

Passive Homing Systems

Passive homing can make use of all the above described seekers, with the possible exception of radar. At the top of the next page a passive homing system is shown in block diagram form. Regardless of the portion



Passive homing system



Semiactive homing system and active homing system

of the frequency spectrum used, the sensor picks up the target and produces a signal representing the placement of the target with respect to the missile. This signal is then fed to the signal converter. The converter changes the signal into the form necessary for comparison with the reference signal from the reference unit.

In the case of conical scanning, as explained in the discussions of basic units on pages 290-291, the converted signal is an AC wave whose frequency is the same as the rotational frequency of the sensor antenna. The signal from the reference unit is made up of sine waves of the same frequency as the signal from the signal converter. These two reference signals are 90° out of phase with each other because the pitch axis is displaced 90° from the yaw axis. The comparators for the two missile axes, in the case of the conical scanning, are both phase comparators.

Other methods of scanning for a target with a passive system can also be employed. The simplest method involves the use of multiple fixed sensors whose pattern of view or detection overlaps along a central axis. When a signal source is so located that all sensors are receiving an equal signal, the central axis of the missile is aligned on the target. Note that there is no necessity for a reference system in this arrangement.

The reference signal would also be altered if another type of scan, such as the rectangular scan used in television practice, were used. The reference would necessarily consist of two sawtooth waves. The period of each of these sawtooth waves would follow the horizontal or vertical scan respectively. For solid coverage of an area, one of the scan periods must be much greater than the other.

The signals from the comparators are fed to their respective channels in the missile auto-pilot.

Semiactive and Active Homing

Semiactive and active homing systems generally make use of the medium of radar. These systems require more equipment than the passive system just discussed. A block diagram of a semiactive homing system is shown in black in the figure at left. An active homing system requires the same components as shown for the semiactive system, plus the transmitter and duplexer which are shown in black.

The signal converter, reference unit, and comparators serve the same functions as explained in the discussion of the passive homing system. Again, there is the variable in circuitry which is necessary to go with the particular scanning method used.

Another special application of an active homing system should be mentioned here to show you the possibilities of homing systems. The application is called beacon homing. The system requires a transponder unit which is planted in the target area. This unit is triggered by radar pulses from the missile at a predetermined frequency. The unit then transmits radar signals back to the missile at the same or a different frequency. The net result is that the missile receives from the transponder unit a return signal which is stronger than its own radar echo signals.

The transponder unit need not be planted on or in the target. The missile homing system, by means of a computer, can use this transponder unit as a navigational aid. Information as to the location of the target with respect to the transponder unit (for example, if a target is two miles northeast of the transponder) must be known, and the information must be preset into the computer of the missile. With this operational setup, there is less opportunity to jam the radar because of the short period of operation.

DOPPLER HOMING. One of the latest applications of radar to an active homing system is the use of the Doppler principles. This phenomenon of frequency change occurs when there is relative motion between an antenna and a target. Doppler homing equipment can be subdivided into two groups of active homers: FM-CW Doppler systems and pulse Doppler systems.

In a *FM-CW Doppler system*, the regular active homing system applies. However, a major difference between the two systems is found in the circuitry and equipment contained in the signal converter block. In a FM-CW system, the frequency of an echo signal has a relationship to the speed of the target with respect to the receiving antenna. With the proper circuitry, this echo signal can

be converted into an indication of the velocity of the target with respect to the missile.

The difference between the frequency of the transmitted signal and the frequency of the returned signal it due to the Doppler shift. The difference can be developed as a beat note between the two frequencies. This beat note or beat frequency is proportional to the velocity of relative motion between the target and radar antenna. A beat frequency of a particular tone signifies that the velocity of this relative motion is of a certain value. A bandpass circuit can effectively be used in the missile equipment so that the equipment can not detect fixed objects or objects moving the same velocity as the missile is going.

An automatically tuned receiver system which sweeps the portion of the Doppler frequencies that are passed on by the filter is used to choose and lock on the target. An automatic frequency control maintains the receiver selection of the desired target. At the closest point of approach to the target, zero Doppler shift occurs because at that instant no relative motion takes place between target and missile. This zero-shift point is used to detonate the missile warhead. A near miss can therefore be transformed into a kill.

This system described here does not have a ranging measurement. Additional circuitry is required to perform this function.

A pulse Doppler system performs the functions of an FM-CW system, but it can also select a target by its range. It also has greater operating range for the same average power output, as is true for any pulsed radar system. For comparison purposes, successive transmitted pulses must be matched perfectly in time sequence both in pulse-timing and RF cycles. This is accomplished in a circuit arrangement known as the coherent pulse Doppler system. In such a system, the transmitter generates short pulses at a repetition frequency which can be continuously varied. The stabilized local oscillator (stalo) furnishes low-intensity priming power to the transmitter to effect phase coherence between successive pulses. A duplexer provides low-impedance paths and effectively reduces the transfer of energy in all but the desired direction. The stalo also provides a local oscillator signal, suitably shifted in frequency, which is mixed with the receiver signals to generate a receiver intermediate frequency. The Doppler receiver, with its associated velocity gate, filters the signal so as to reject the undesired Doppler frequencies.

Now that you're familiar with homing systems, let's consider the paths that missiles follow when the homing systems are in operation.

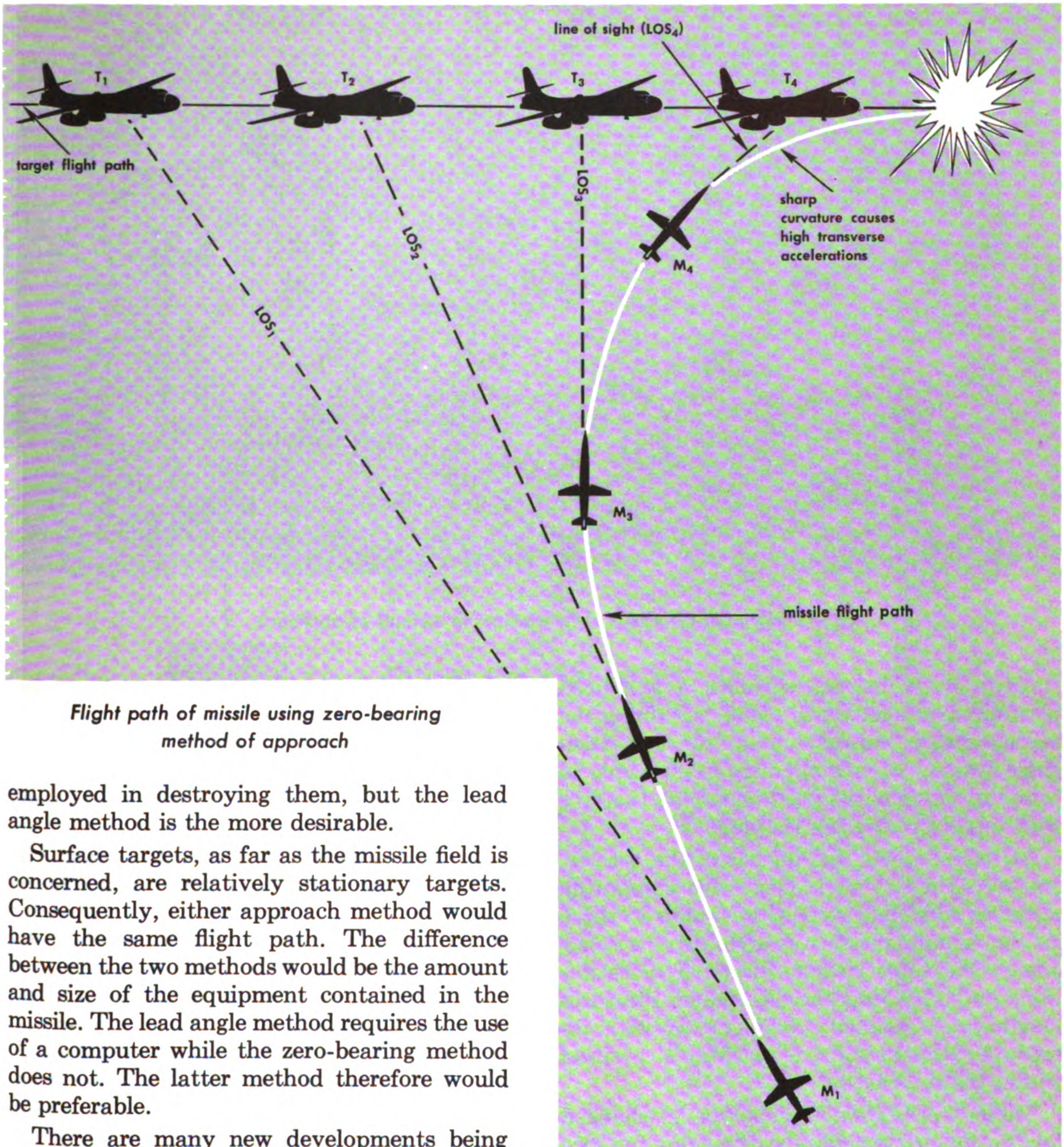
Methods of Missile Approach to Target

There are two methods by which a homing missile can approach a target. You were introduced to these methods in the chapter on missile trajectories. The first method consists of flying directly toward the target at all times; this method is known as *zero bearing*. A possible flight path of such a method is shown in the illustration on the right. The illustration shows a situation in which the target is moving. You can easily visualize that a flight path to a stationary target would be a straight line. Notice the curvature of the flight path as the missile approaches the path of the moving aircraft. The sharp curvature sets up tremendous lateral accelerations. These transverse accelerations represent a strong objection to the use of the zero-bearing method against high-speed targets. Another objection to this method is that the speed of the missile must be extremely fast compared to the speed of the target.

The other method of approach is called *lead angle*, sometimes referred to as *constant true bearing* or *collision course*. This method requires the addition of a computer to the system. The computer continually predicts the point of impact. If the target takes no evasive action, the point of impact remains the same. If the target does take evasive action, the computer sends signals to the autopilot of the missile to correct for a new point of contact.

Using the lead angle method, the missile must still travel faster than the target, but not as fast proportionally as with the zero-bearing method of approach. Also, because of the slower speed and the smaller and less rapid changes in attitude, the transverse accelerations of the missile are comparatively small.

Airborne targets are moving targets. Either of the two methods of approach can be



Flight path of missile using zero-bearing method of approach

employed in destroying them, but the lead angle method is the more desirable.

Surface targets, as far as the missile field is concerned, are relatively stationary targets. Consequently, either approach method would have the same flight path. The difference between the two methods would be the amount and size of the equipment contained in the missile. The lead angle method requires the use of a computer while the zero-bearing method does not. The latter method therefore would be preferable.

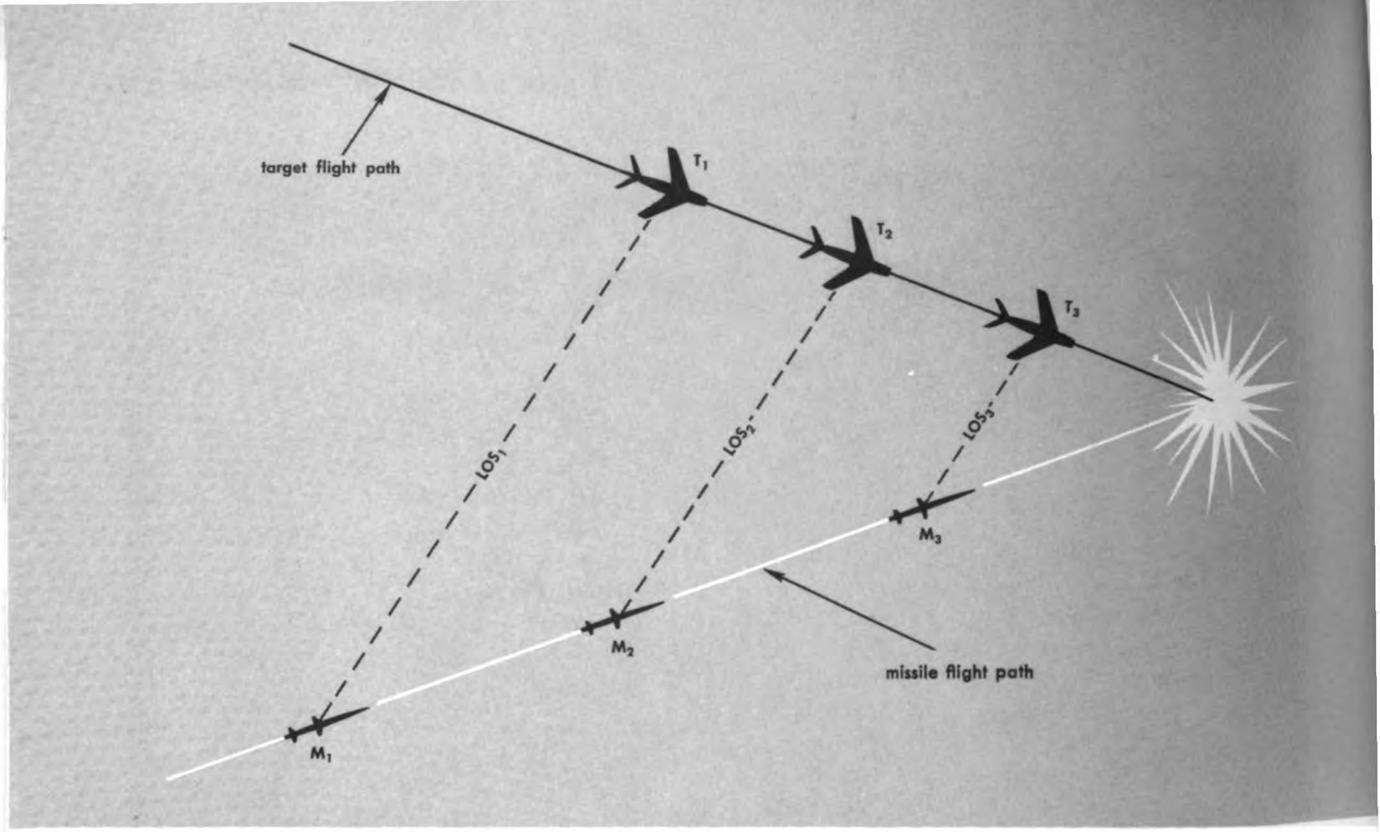
There are many new developments being made in the field of homing systems. Research and development will continue to have an important part in the construction of newer and more acceptable homing systems.

TERMINAL INERTIAL SYSTEMS

There are two specific terminal inertial guidance systems. They are known as the *constant-dive-angle system* and the *zero-lift (ballistic) inertial system.*

Constant-Dive-Angle System

On page 455 is shown a general block diagram of a constant-dive-angle system. As in any terminal flight system, midcourse guidance equipment directs the missile to some point in space known as the release point. At this point the midcourse guidance equipment is disabled, and the terminal guidance system takes over control of the missile.



Flight path of missile using lead-angle method of approach

As in other inertial systems, a stabilized platform is the reference plane containing the accelerometer sensors for a constant-dive-angle system. Missile circuitry is then able to compute the missile's position during the dive with respect to the release point.

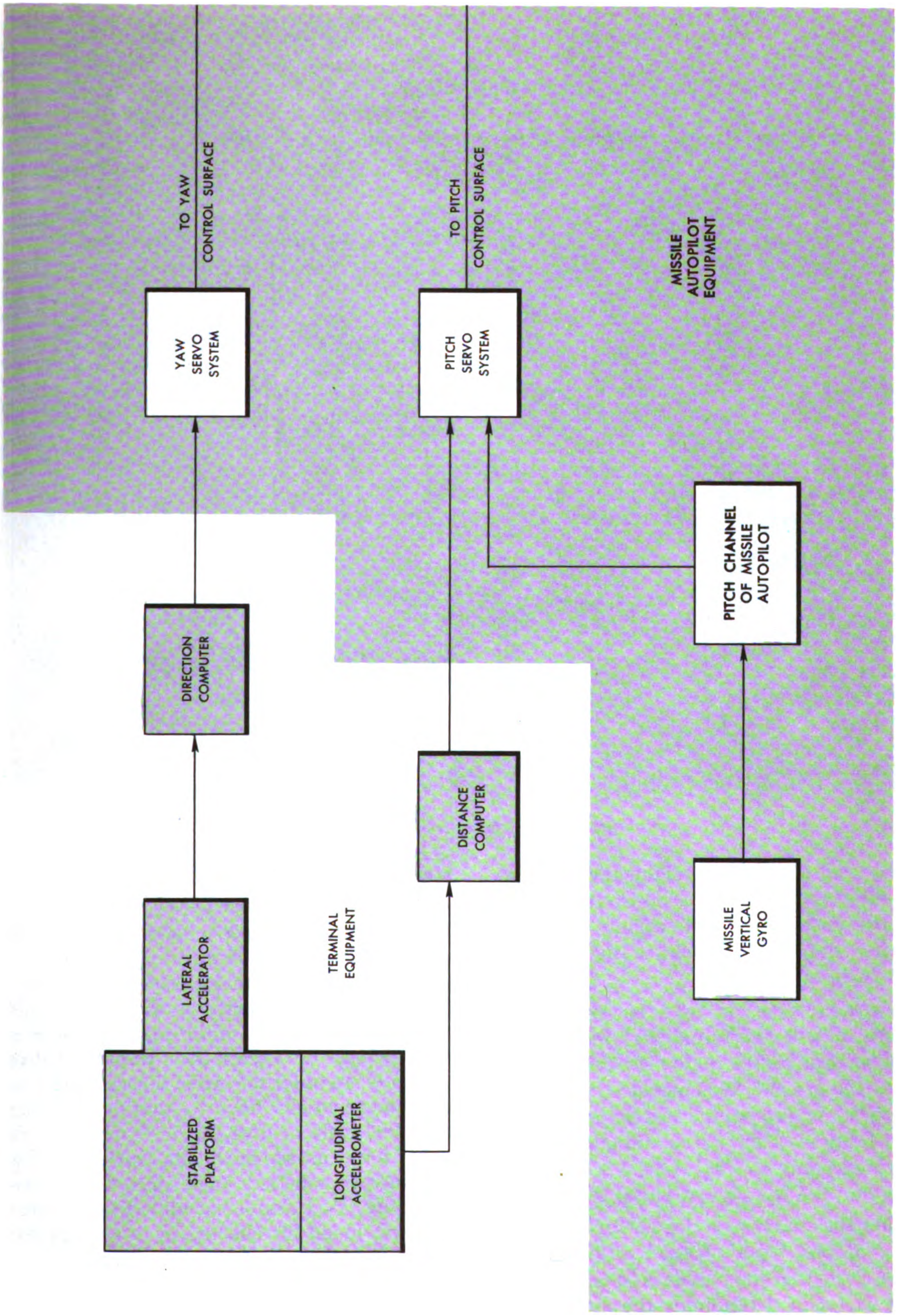
At release point, the output of the accelerometers is fed to their respective computing channels which are somewhat similar to those in the simple inertial system. One accelerometer measures lateral (direction) acceleration, and the other accelerometer measures longitudinal (distance) acceleration with respect to the stabilized platform's position within the missile.

These acceleration signals can be changed by the integrators within the computer into velocity signals and then into a signal representing position. However, the distance computer channel generally does not need an output representing the position error. The computer channel for the distance channel of the constant dive angle has only one integrator, and the velocity signal is sent to the pitch servo.

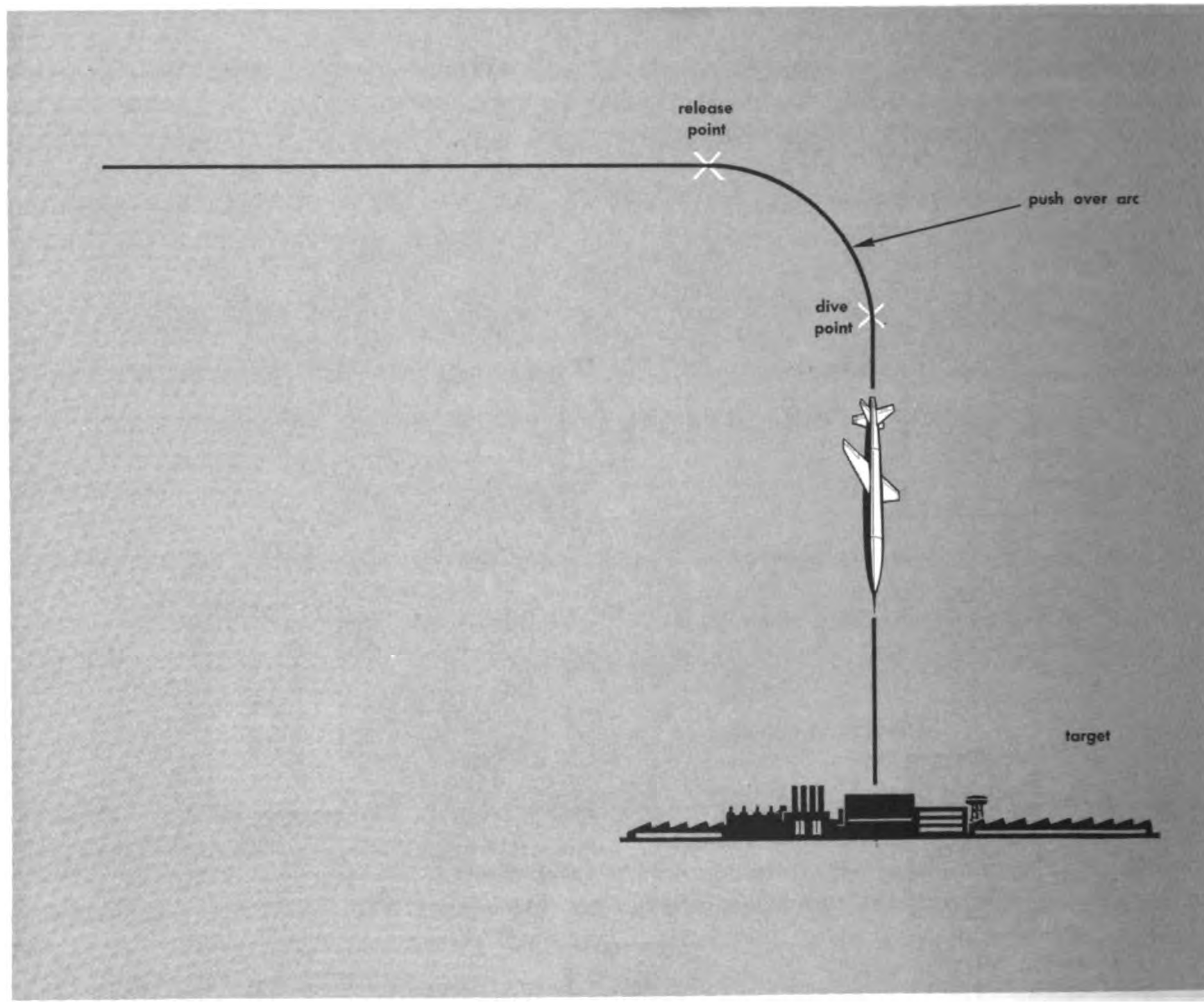
When the velocity signal is of the correct constant value, there is no output from the computer to the pitch servo. The method of accomplishing this error-only output was explained in the preceding section in the discussion of simple inertial systems. If the computer output should differ from the desired value, there is an error output at the pitch servo. The error output causes a control action that corrects the dive angle to the proper amount.

Vertical-dive system. A variation of the constant-dive-angle system is a vertical-dive system. This system makes use of the distance computer which has two integrating blocks so that position error is the signal output. Therefore, the desired signal output of the computer channels is a zero-error signal signifying that the missile in on the vertical path.

The illustrations on pages 456 and 457 show the flight path of vertical-dive and constant-dive-angle systems. They make noticeable the fact that the vertical-dive is a special condition of the constant-dive-angle system. After the missile has completed its push-



Constant-dive-angle system



Flight path of a vertical-dive integral system

over arc and is diving straight toward the earth, any acceleration in a horizontal plane causes the missile's control surfaces to react and bring the missile back to its original vertical path.

PUSH-OVER ARC. You may be wondering how the missile performs its movement through an arc before taking a straight-line course to the ground. This movement, known as push-over arc, is accomplished by precessing the vertical gyro of the automatic pilot about the pitch axis only.

Naturally, there are many factors that determine the number of degrees of arc and the rate of precession of the vertical gyro. The dive angle that the missile is to take is the primary factor in determining the number of degrees of vertical precession. Another factor is the angle of incidence of the wings.

Looking at the upper missile in the accompanying illustration of missile dive attitude, you can see that, if a missile's longitudinal axis were pointing straight down, the wings would have some lift because of the angle of incidence of the wings; thus the missile would not dive vertically. In order to correct for this, the missile would actually have to nose over farther than the angle of dive in the manner shown by the lower missile in the illustration. This is true of any constant angle dive. This compensating effect also is calculated to do away with the added lift effected by the wings due to the fact that a missile, as a general rule, travels faster in the dive causing the wings to exert more lift.

The rate of precession, or how fast the missile is to nose over, is an engineering problem involving, among other things, the aerodynamic characteristics of the missile.

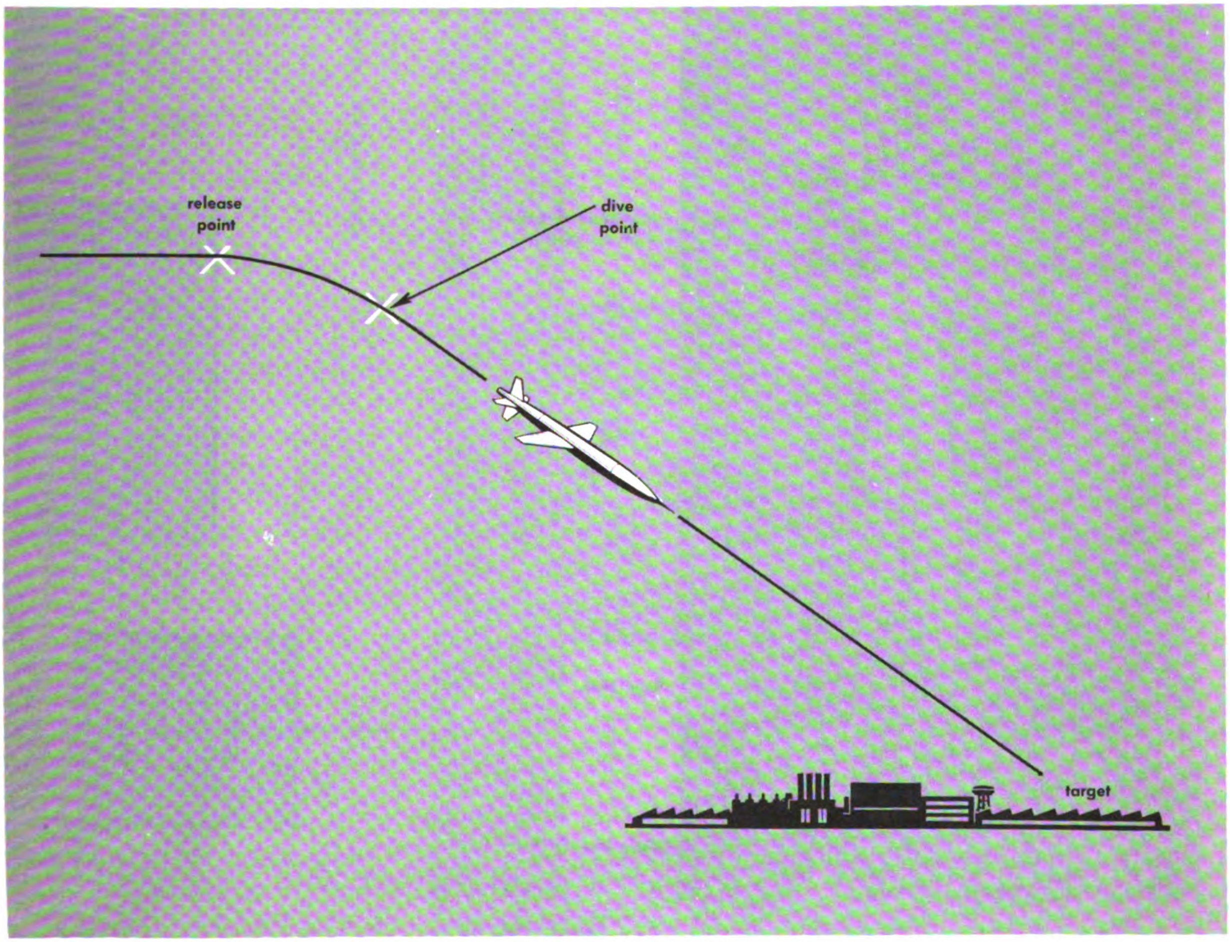
When this arc through which the missile moves is completed, a point is reached called *dive point*. At dive point the automatic pilot is cut off from the yaw and pitch servos and has no further effect on the control surfaces.

Zero-Lift Inertial System

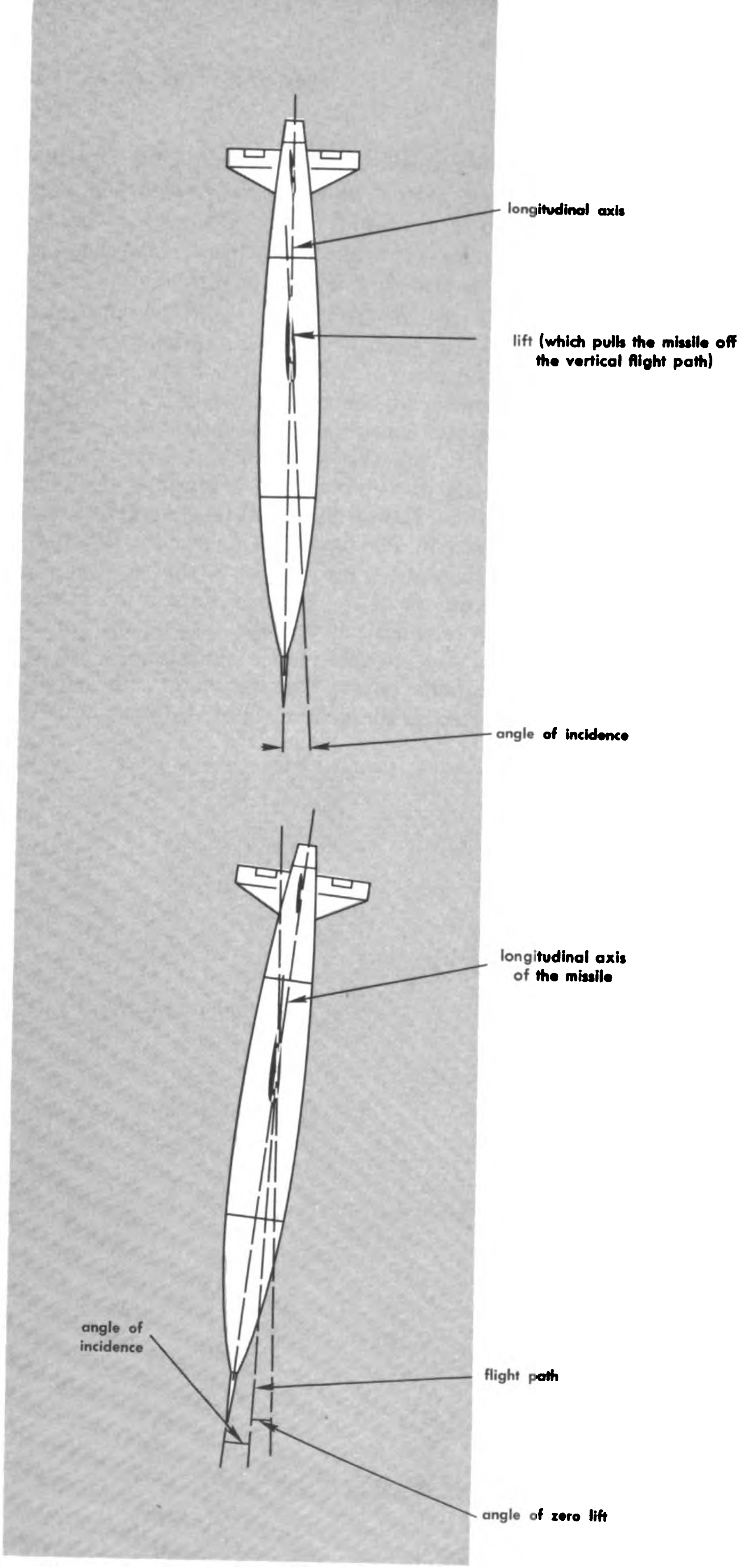
A typical zero-lift inertial system is shown in the diagram on page 459. The diagram also shows the relationship between the terminal system and the missile control system (the autopilot).

The terminal equipment can be broken down into two separate functions. The first function is to establish the flight path, or programmed path, by means of a constant-speed motor whose rotor moves the wiper arm of a potentiometer. The second function is to keep the missile on this programmed path by means of an accelerometer.

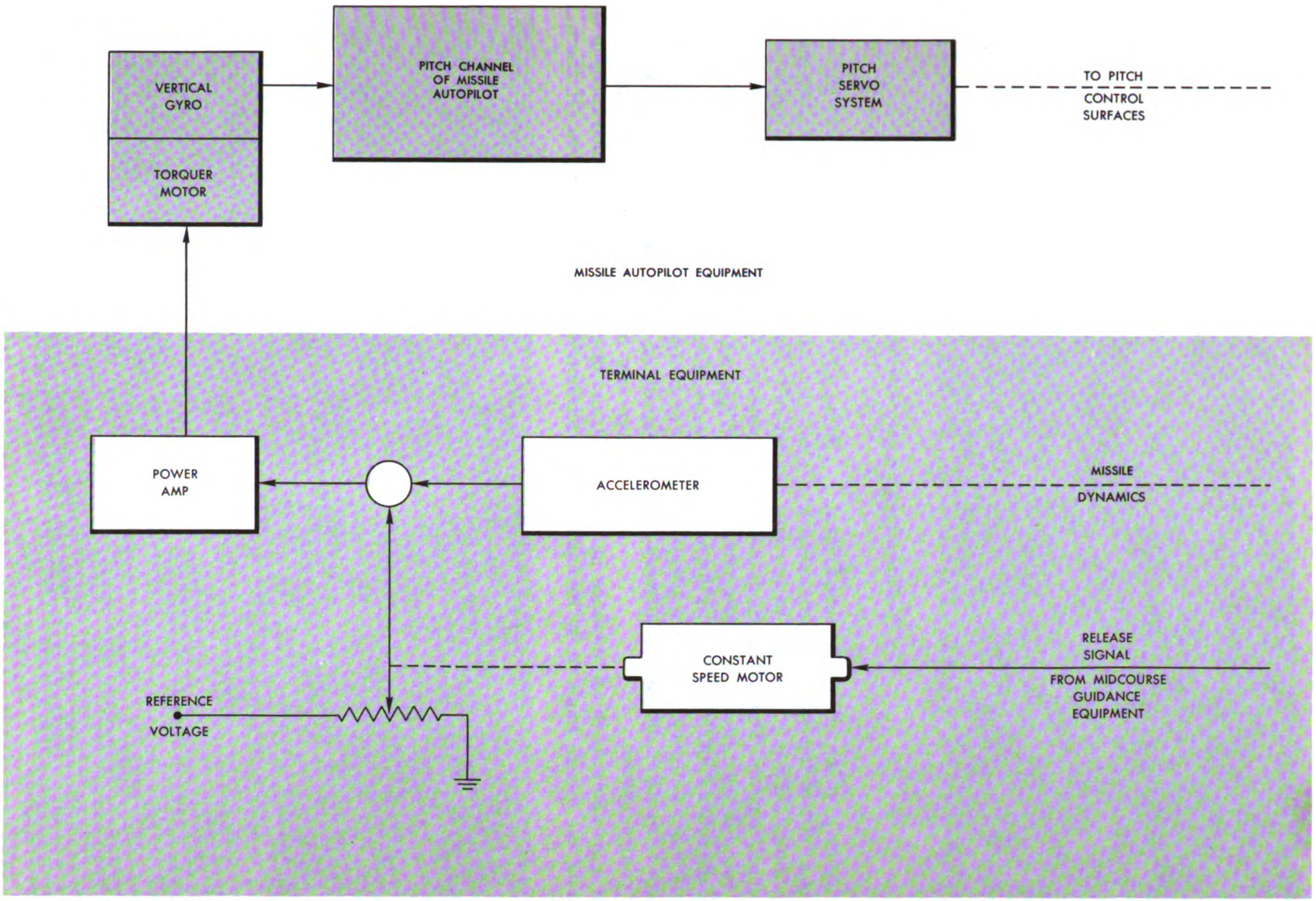
If the desired path is to be theoretical ballistic path, which is a parabolic curve, the wiper arm of the potentiometer has to be moved from the ground end of the potentiometer at a constant rate of speed. If the voltage taken from the wiper arm (the programmed signal) were plotted on a graph with respect to time, the result would be a straight line, as shown in the figure on page 460. When this voltage is fed into a motor, the resultant displacement of the motor's rotor is an integration of the input voltage. As you will remember, the integral of a constant-slope line is a parabolic curve. The graph of this curve is plotted in the bottom figure on page 460.



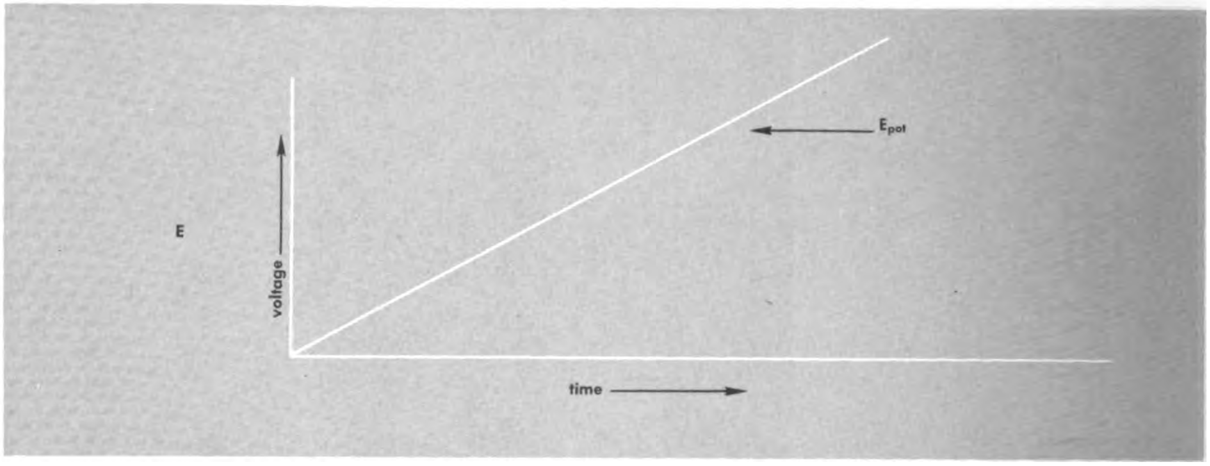
Flight path of a constant-dive-angle inertial system



Dive attitude



Zero-lift inertial system

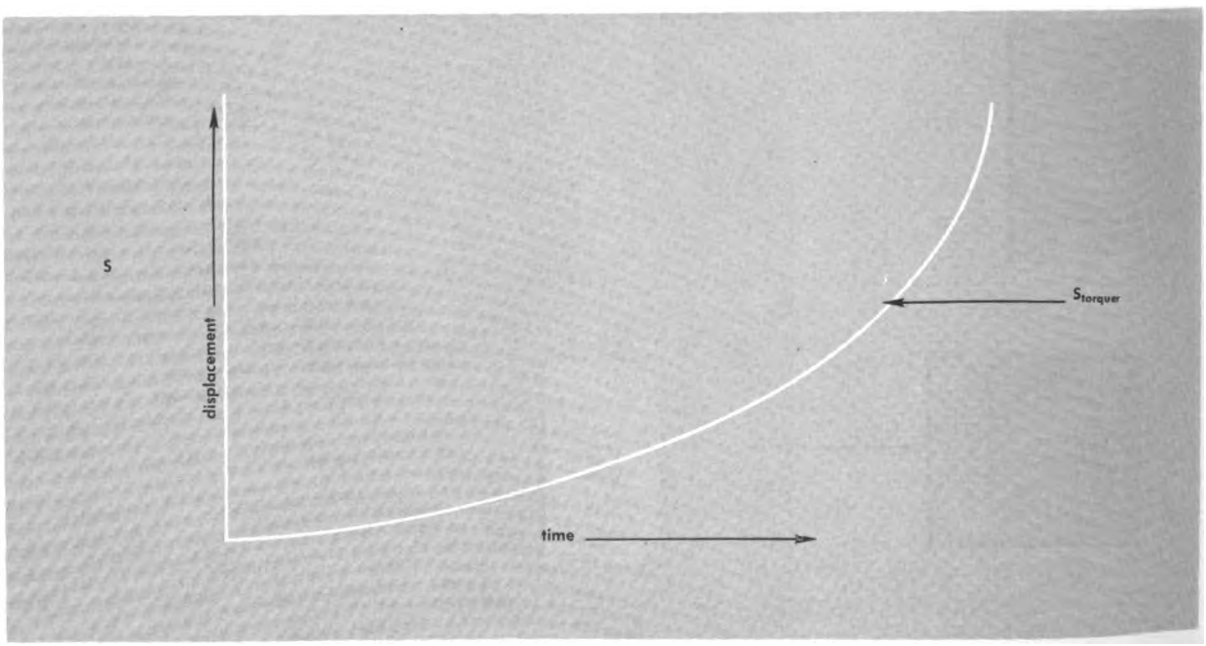


Voltage taken off a potentiometer

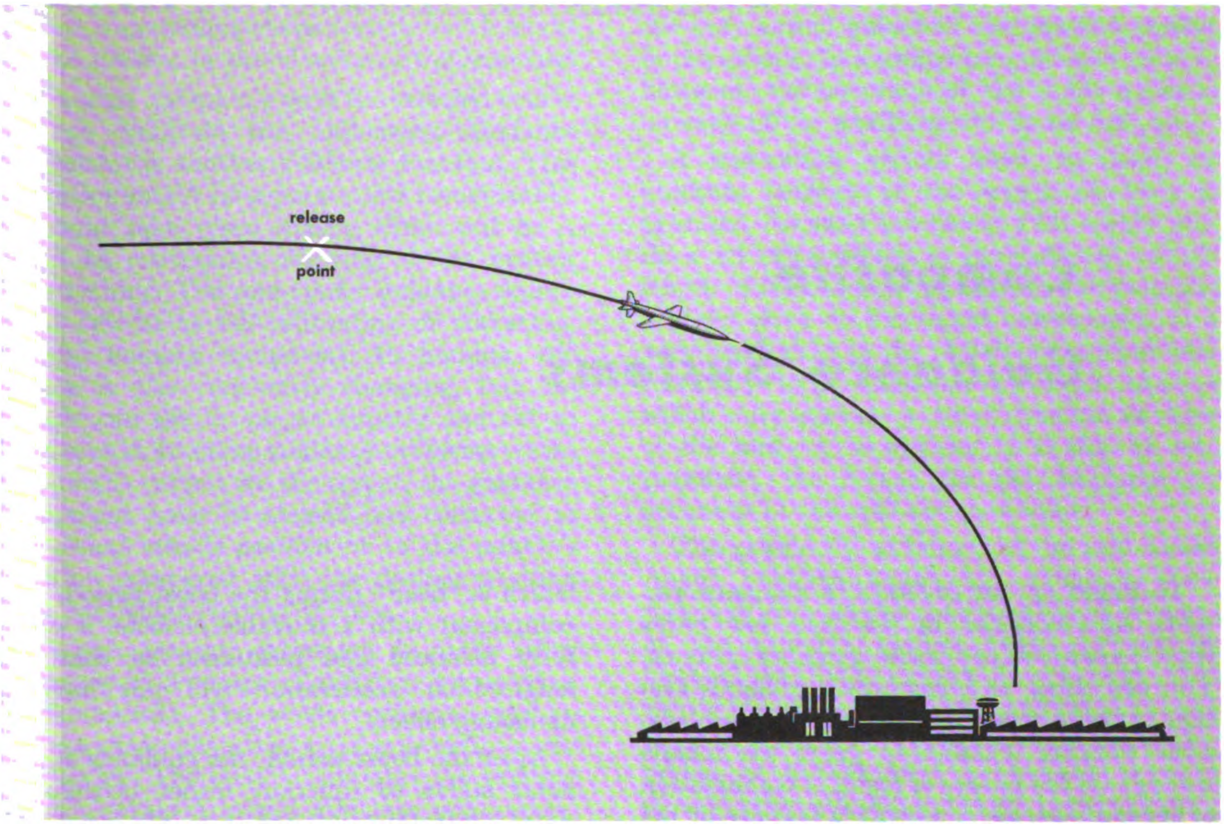
The motor to which this programmed signal is fed is a part of the vertical gyro, called a torquer motor, and is not shown in the preceding block diagram. The displacement of this motor rotor is used to precess the vertical gyro of the missile control system in the *pitch* axis; therefore, the gyro presents a parabolic path as a *reference* to the pitch channel of the missile control system.

With a parabolic path as reference in the pitch axis, the missile tends to fly that path. Because of the angle of attack of the wings and the thrust of the engine, the missile actually flies some other path, unless some compensation is made for these factors.

This compensation is accomplished by means of an accelerometer which is mounted so as to be sensitive to accelerations in the vertical of the missile. Thus, if the wings have any lift, the accelerometer senses this lift and originates a signal which corrects the precession of the vertical gyro. If there is



Displacement of torquer-motor rotor



Flight path of a zero-lift inertial system

some lift present, the signal from the accelerometer adds to the programmed signal in the mixer and precesses the gyro at a faster rate. If the missile noses over too far and *negative* lift is present, the accelerometer originates a signal which subtracts from the programmed signal in the mixer and slows up the precession rate of the gyro. The missile therefore flies a predetermined (theoretical ballistic) path to the ground as shown in the illustration above. The term zero-lift is applied to this system because the accelerometer signal compensates for any lift in the vertical axis of the missile.

Midcourse guidance brings a missile to the desired release point, and it also sets up the circuitry for the terminal guidance equipment, including the release signal to drive the constant-speed motor. The purpose of the power amplifier in this diagram is, as its name implies, to take the input signal and amplify its power so that it can drive the torquer motor.

In the case of either inertial system, the accuracy of the terminal guidance equipment cannot *exceed* the accuracy of the midcourse guidance in determining release point. Of course, the terminal inertial systems themselves have inaccuracies which require compensation and further research.

At this point in the text you have studied the systems that guide a missile from the time it is launched to the time it makes contact with its target. By now, you should be well aware that because of the nature of a missile flight, a single guidance system would be impractical, if not impossible, for carrying a missile through all the phases of flight. Because of this condition, a combination of guidance systems is used in any one missile. The next chapter elaborates further on the subject of composite guidance systems.

Composite Guidance Systems

To round up our coverage of guidance systems, let's consider in the next few pages what constitutes composite guidance systems. You can think of a composite guidance system as being several individual guidance systems working together to guide the missile through all the phases of flight. A combination of several systems is necessary because of the wide difference in requirements that have to be met. During the immediate launch or boost period, extreme acceleration prevents the use of normal guidance components. Such accelerations act to close relays, precess gyros, and saturate accelerometers which have the required sensitivities for *ordinary* flight conditions.

If a standard midcourse guidance system were to be used during launch as well as during the midcourse phase, it would have to be modified to withstand the launch acceleration. The modification must be on an alternate set of relatively insensitive guidance components or a temporary alteration of the regular components.

Usually, regular guidance system components that contain movable parts are locked in position or, at least, the circuits into which they work are neutralized to withstand the strain of launching operations. The missile airframes are designed sufficiently stable so that control settings made before launch will enable the missile to make a stable flight throughout the boost period. Even in a system which is "dead" during the boost period, any component with a movable element is carefully balanced and positioned so that the longitudinal acceleration will not have a damaging effect.

After the boost period, the guidance equipment starts to function, either because of timer operations or the loss of the boost acceleration. Midcourse guidance is not always put into operation at this time. In ground launched missiles the climb out period is necessary to get up to the proper altitude and to get in the proper position for midcourse guidance to take over. The climb out period can be supervised either by a preset guidance system or by a form of remote control. In some missiles there is no climb out period as a separate part of guidance. In these missiles the midcourse system takes over immediately following launch.

The mid-course trajectory usually is calculated to start at some point remote from the launching area. There are several reasons for this. Since mid-course trajectory follows a great circle path or some similar defined characteristic, a launching on the trajectory would destroy the security of the launching location. By plotting the missile's course and searching along its path, an enemy could locate the launching site. There also can be technical reasons for delaying the start of the mid-course trajectory. A mid-course system using some terrestrial feature, such as

magnetism or mapmatching, requires a particular, outstanding characteristic which might not be present in the vicinity of the launcher.

To align the missile with the selected course and to calibrate the mid-course guidance equipment is another guidance function. This operation might be separate and recognizable. It could take place during climb out or it could be continuous, starting at the instant of launching. The mid-course guidance equipment can be as self-calibrating or correcting as any homing equipment. In homing, the course is also constantly changing, but the missile can see its target directly and act to intercept it. In other guidance problems where the missile cannot see the target, the guidance equipment must be aligned and corrected to the *actual missile position* or to the missile's position compared with the desired position at a particular instant of time.

In one system, correction is made after reference to a source of information external to the missile. The whole correction system may be external such as a radar or radio tracking device. The device computes the existing error in terms of missile guidance parameters. Then it codes them and transmits them to the missile on a command link. In another system, the correction is contained in the missile. Such a system, for example, uses Doppler radar to measure ground velocities in transverse and longitudinal directions, and it also uses air speed indicators or fixed beacon plotting radar.

When a missile "brain" decides that sufficient, accurate information is available as to the preset position, the mid-course guidance is put into action. Possibly the mid-course guidance system has already been functioning but has not been the primary source of error information. The signals of the mid-course guidance and the alignment systems have been mixed in the missile "brain" and the accurate error calculated. The mid-course guidance then takes control, having been accurately and dynamically balanced under the conditions of the actual flight.

In many missiles, the mid-course guidance system is considered the dominant one, with other guidance systems relegated to the position of auxiliaries. The mid-course system operates over the longest range and time. In short-range missiles, the mid-course is often the only actual guidance equipment installed in the missile. In a composite guidance system, however, the accuracy of the flight depends on each of the auxiliaries to do its full job.

A mid-course system must, in itself, be accurate because of the length of time it operates. An error-producing drift would produce an error which is proportional to the length of time it is accumulated. For example, a given error at the target corresponds to a very small drift in the mid-course system and relatively large drifts in the other systems which only operate a small fraction of time of the mid-course system. It is easy to see, therefore, the reason for the emphasis on accuracy in the mid-course system.

The possibility of failure of the mid-course (main) guidance system also exists, so some means of providing a guidance signal in spite of the primary system failure is usually incorporated in a missile. The system that serves this purpose is a true auxiliary. It may be used during the correction and alignment period, or it may exist for the sole purpose of a standby guidance system. This auxiliary might simply be an arrangement set to take over if primary signals are lost, or it may provide for inertial continuation of the last correct signal conditions. As stated previously, it could be a completely separate system. Standby guidance is so arranged as to return the guidance function to the primary system if the primary system begins to function properly again.

CONTROL MATRIX

The selection of the guidance system which can control the path of the missile is a function of a special system. It could be that guidance-system selection circuits are scattered around within the electronic equipment of the missile. Each *system-enabling* circuitry is located in close proximity to the particular system it causes to function. If all the selection circuits are organized into one block, the circuitry is called the control matrix. Separated or unified, the functions of the matrix components are the same. Their functions are to select a guidance system operation in the proper sequence for the flight.

Control of a missile during flight can be provided for by more than one source. Some signal is set up to designate the phase of flight the missile is going through. This signal may come from playback equipment which uses a recorded tape to determine phases, or it may come from a combination of signals from various units such as radio control or primary guidance. A control matrix acts upon the signal whatever its source, to provide the command which switches in the correct guidance function for the particular phase or period of flight.

A control matrix can be likened to an automatic telephone switchboard, automatically transfering the correct signal to the control system regardless of conditions. By being able to provide a guidance function in spite of any eventuality, the control matrix furnishes the missile with the most correct guidance function.

If the mid-course guidance should become inoperative or unusable, the matrix switches in an auxiliary guidance system to hold the missile on course with an accuracy only secondary to that of the mid-course system. If the primary mid-course system again becomes operative, the control matrix returns the guidance function to it. A schematic drawing of a circuit which represents matrix functioning is possible through the use of Boolean notation.

In Boolean notation, each signal or operation is coded as a letter. The letter alone means that the signal is present. A letter followed by a prime (′) is an indication of no signal or off condition.

An equation represents the combination of signals required to effect a certain output. If the combined signals are separated by dots, the dots can be replaced by the word *and*. If the terms are separated by a "+" or "V," the sign is a substitute for the word *or*. Grouping the terms in parentheses implies the same operation as ordinary algebra. The signal represented outside the parentheses must be present with whatever combination is represented within the parentheses.

The matrix, then, can be considered as the automatic guidance switchboard or guidance sequencing computer. Inasmuch as the signals are of the on-off variety, they can be considered as binary digits. A succession of these binary digits results in a signal, the nature of which can be considered to be dyadic (consisting of two elements). The control matrix operation is analogous to a digital computer, and for a combination of digits fed in, the proper digit comes out.

VELOCITY-DAMPING DOPPLER RADAR

A Doppler radar used for velocity damping of an inertial system is somewhat different from the homing type of Doppler. The velocity-damping Doppler requires antennas to measure the forward and lateral components of velocity. These antennas have to be direction stabilized so that the velocities along, and normal to, the flight path can be measured to nullify errors that a drift angle creates.

The antenna mounting has two antennas looking down and forward at a slight angle away from the roll axis of the missile. A third antenna looks to the rear. A comparison between the signals from the two front antennas is used to align the direction of the missile. A comparison between the forward and rear antenna signals gives the forward velocity Drift is computed from the angle between the antennas' fore and aft axis and aircraft heading.

A Doppler radar similar to the equipment used in bomb navigation equipment is used with a special marriage unit employed to couple the Doppler output into the missile computer.

Doppler would not be used continually on a flight, for such use would violate security. On long, over-water flights, Doppler is of little use because insufficient return is received from a relatively smooth, flat surface such as an ocean.

As stated previously, this section ends the manual's treatment of missile guidance systems. At this point you should have an understanding of the methods and means of missile control and guidance. You should also know the principles involved in missile flight. But just how is a missile used as a weapon of war? The following chapter answers that question.

Chapter 10

Guided Missile Tactics

With the advent of guided missiles, two extremes of thought have come into existence relative to the nature of future wars. Probably the more common is the push-button concept. This concept envisions future wars being won or lost by intercontinental battles with supersonic long-range missiles carrying A-bomb or H-bomb warheads. Thus, there would be little or no ground fighting involving the use of foot troops, small arms, tanks, and the like.

The other concept is that atomic and hydrogen bombs are an improvement over the bombs of the past and that guided missiles are merely more effective carriers than the bombers of World War II. Therefore, advocates of this latter concept contend that the general mode of warfare will not be altered. War must still be won on the ground, according to them, and all weapons exist to help the ground soldier advance toward a specific objective. This concept reasons that even if decisive military action shifts to missile weapons, ground troops must repel enemy invasion attempts, seize bases needed for launching missiles, and physically occupy critical territory to prevent the resumption of hostilities.

Both of these concepts are subject to considerable discussion. In any event, however, missiles are certain to play an important role in the isolation of the battle-field and neutralization of important targets.

SURFACE-TARGET SELECTION

Long-range missile attacks are designed to bring about, through the systematic application of force to a selected series of vital targets, the progressive destruction and disintegration of the enemy's war-making capacity to a point at which he no longer retains the ability or will to wage war. The part played by missiles in attaining this goal consists, most of the time, of surface-to-surface and air-to-surface missiles employed against selected surface targets.

Considerations given to the selection of surface targets for missiles differ little from the considerations given proposed surface targets

of conventional aircraft during World War II. Targets remain essentially the same. Guided missiles, therefore, are needed not for attacking new ultra-modern ground targets but the same long-existing strategic targets of the past which are highly vulnerable from the air.

Surface-Target Classification

Generally, surface targets are classified as being either tactical or strategic. Tactical targets are those which have a direct influence on the course of a battle in progress. For instance, military personnel, tanks, and gun emplacements in the battle zone are tactical targets. Such targets usually appear directly at the point of enemy contact, the front lines. They are generally temporary, transient, or fleeting in nature, and they require attack by weapons which can reach them in a minimum of time. Other tactical targets are scattered over a front and extend deep into enemy territory. An enemy launching site may require immediate attack even though a hundred miles from the battle lines.

Strategic targets are normally located well beyond the battle area, and their destruction does not exhibit a direct or immediate influence on the course of a particular battle. These targets are usually located deep within the enemy nation. They are generally large installations of great importance to the productive power of the nation. Industrial plants, mines, and oil fields are examples of strategic targets.

With the development and employment of guided missiles, the tactical target area moves closer to the enemy home front. If troops in a battle zone are classed as a tactical target, then surely troops hundreds of miles away who are loading aircraft to be immediately thrown into that battle must also be tactical targets. Thus, it becomes increasingly difficult to differentiate between tactical and strategic target areas on the basis of distance from the line of battle. This chapter, therefore, presents the basic considerations given to the selection of missile surface targets with no major attempt to differentiate between tactical and strategic situations.

Specific Target Systems

The ultimate objective of attacks on surface targets is to weaken and immobilize enemy troops so that they are more vulnerable to destruction and defeat. The major targets for consideration are industrial centers in the heart of enemy territory. Successful attacks on such centers not only affect civilian morale but drastically affect the flow of supplies and war implements to the enemy troops. The enemy's power to make war must be destroyed. This means that his factories, communication systems, food stockpiles, and fuel and oil supplies must be destroyed, and that his people and the places where they carry on their daily living must be destroyed.

Generally speaking, the productive capacity of a nation is dependent on its strength in four broad categories:

a. Raw Materials.

b. Processing and end-products manufacture.

c. Supporting services and production of basic equipment.

d. Imports.

Each of these groups is a source of targets. The relative importance of these targets depends on an analysis of the specific objectives to be attained and the war situation at the moment.

The first group contains the mines, forests, oil fields, and farms from which are extracted the nation's raw materials.

The processing and end-products group pertains to the factories which convert the raw materials into finished products for civilian as well as military use. Aircraft plants, armament plants, and refineries are a few of the many targets in this group.

The supporting services and basic equipment group consists of such supporting facilities as power plants, internal transportation systems, and research and development agencies. Constituents of this group are not directly engaged in the production of war weapons, but they support the producing units.

A nation's imports are also of major importance during time of war. Targets in this category consist mainly of seagoing vessels, railway cars, and pipelines carrying critical materials to the enemy nation. Many

targets of the imports group are most vulnerable while in transit. However, their destruction under any conditions produces favorable results.

Target Considerations

Some of the criteria which should be considered in the selection of surface targets are discussed below. Keep in mind that the foremost characteristic of any missile target is that it should be beyond the capability of a more conventional and less expensive weapon.

IMPORTANCE OF TARGET TO ENEMY. A target should be of such importance that its destruction would deprive the enemy of a means of supplying critical materials necessary for carrying on the war or a particular campaign. The importance of certain targets may vary with the progress of the war or theater of operation; that is, at one stage in a war, or in a particular theater of operation, cutting off the nation's imports may be of major importance. At some later date, or in a different theater, destruction of oil refineries may receive highest priority. In any case, the importance of the target should justify the use of missile. Also, the service or products of the target should be of a nature for which substitution or elimination from use is not feasible.

RESERVE CUSHION MAINTAINED BY THE ENEMY. The target selected should be one which has no large reserve production capacity nor alternate sources of supply. Among other things, reserve production capacity refers to idle facilities that may be put into operation on short notice. Too, a certain amount of cushion exists where facilities are being used for the production of low-priority war materials, but which can be easily converted to high-priority war production by making the necessary conversions, replacements, and expansions in equipment and operation. Additional cushion exists in plants and materials used in the production of commodities for civilian consumption. The use of many such plants and materials could easily be restricted to military production. In many cases, the amount of cushion may determine how much of the overall industry must be destroyed in order to effectively retard the flow of finished products from that particular industry.

DEPTH TO WHICH TARGET DESTRUCTION WILL BE FELT. Depth refers to how soon and to what extent an attack on a particular target will impede the enemy's war effort. An attack on a target engaged in the manufacture of component parts of some war weapon should result in a critical reduction in the output of finished weapons. The products of the target under consideration should not be available in large enough stockpiles to permit uninterrupted production of finished items even though output from the target has been delayed.

VULNERABILITY OF THE TARGET. A target under consideration must have physical characteristics which will allow damage to be achieved by the attacking missile. A highly vulnerable target can be damaged effectively with a minimum expenditure of missiles. Consider, for instance, an aircraft landing strip, or an electric power plant. Both of these targets are highly vulnerable since a single missile hit might easily render either of them unusable. On the other hand, a target located well underground and housed within reinforced concrete would be much less vulnerable. Such targets would require the use of a larger number of missiles of greater destructive power, and even then damage may not be sufficient to completely halt all operations.

RECUPERATIVE POWERS. Recuperative powers pertain to how quickly and easily the damage to a target may be repaired and the installation returned to normal operational status. In the case of the landing strip previously mentioned, filling of craters would normally place it back in operation. The electric power plant, however, has low recuperability in that minor damage to its equipment may keep it inoperative for a long period of time. Repair is more complicated and time-consuming than in the case of the landing strip. Also, factories and other agencies relying solely on this plant for power will be inoperative during the recuperation period. A target having a high recuperability rate would necessitate frequent missile attacks while a target of low recuperability would require attacks at longer time intervals.

There are cases in which attacks on a target can be planned to allow for a high rate of

recuperability following occupation by our military forces who may desire its use. An example of this is the temporary denial of use of an area or installation due to a nonpersistent chemical agent. After a certain minimum time, our forces could enter the area without harmful effects.

DISPERSAL OF TARGET-SYSTEM COMPONENTS. A target system's components should be concentrated in few enough places that a reasonable number of attacks may destroy it. It is important to know the percentage distribution of production among plants producing a major enemy weapon when those plants are distantly separated. Plants producing the greater percentage of the nation's total output of that weapon would no doubt be given top priority. On the other hand, if the component parts of some weapon are produced separately by a number of widely dispersed plants, then the plants producing the most important parts would probably be given first priority for attacks.

IDENTIFICATION OF THE TARGET. The ability of the attacking missile guidance system to identify a target is of utmost importance. Identification can be acheived more easily if the principal elements of the target exhibit characteristics that make the target easily identifiable. For example, targets emitting heat, light, or some other form of energy are ideal for missiles utilizing that respective guidance equipment.

ENEMY OPPOSITION TO ATTACK ON THE TARGET. The enemy's defense of a target is an important consideration in determining its suitability for attack. Target defense quite often indicates the importance of that target to the enemy. Both the ground and air defense of the target should be evaluated if such information is available. The probability of the enemy employing countermeasures in the form of aircraft fire, antiaircraft fire, interceptor missiles, decoying homing guidance systems, and jamming electronic guidance systems must be considered. The elements of surprise, supersonic speed, and jam-proof guidance systems greatly decrease the possibility of effective enemy opposition.

SIZE AND LOCATION OF THE TARGET. A target must have an area commensurate with the accuracy of the missile used, and the principal elements of the target must lie within the range of the missile. Suppose a preset surface-to-surface missile capable of 150 miles range has a standard deviation of 0.1% of the distance from launching site to the target. This means that at a target range of 100 miles the missile may strike the surface approximately 528 feet from the calculated center of the target. If the target is large enough, the center of dispersion of the missile's warhead will still be somewhere on the target. On the other hand, if the target is small relative to the accuracy of the missile, a complete miss may result.

There is the possibility that the accuracy by which a target can be located is not comparable with the accuracy of a particular missile guidance system. Consider a preset missile which has an accuracy of 200 feet deviation at a range of 100 miles. Such accuracy is of no value when the target cannot be located with respect to the launching site with an accuracy better than 1500 feet. The Germans encountered this problem when trying to locate the center of London with respect to their missile launching sites. Thus, target location in some cases may be a greater limitation than accuracy of the missile.

Within a battle zone, targets are usually detected and located by direct observation or map data. But in the case of long-range targets in enemy-held territory, such means are difficult to employ. Chances of survival of the observer are highly questionable, and existing map data may be inadequate. In the first place, the only accurate representation of the earth's surface is a globe, but a globe is impractical for aeronautical planning. This spherical surface must be represented on a flat surface. *Projections* of this type introduce distortion in geographical coordinates. Various methods of projection are used in an attempt to keep distortion at a minimum and to make the projection most useful for the purpose it must serve.

The role of military intelligence is to detect and accurately locate targets with respect to existing maps. Various sources of information must be utilized to determine *where*, *how*, and *how hard* a target must be hit. Intelligence

must also keep abreast of the appearance of new targets and changes that may occur in already existing targets. Sources of information for these and other considerations of enemy targets may be observations from piloted aircraft, observations made by agents in enemy territory, or interrogation of prisoners of war.

Another method of improving existing map data and accuracy of missile target location is aerial mapping. Aerial mapping is done by piloted aircraft or by reconnaissance missiles controlled by radar triangulation stations. The missiles fly over the target area taking vertical photographs. The friendly radar tracking the missile provides desired survey control. Accurate map data is obtained by synchronizing each photograph with the range and azimuth from the radar station to the missile at the time the photograph was taken.

Successful interpretation of aerial photographs depends on the skill and experience of the individual interpreter, the reference material available to him, the quality of the photographs, and the time available for the interpretation. The most capable interpreters in such fields as oil production, steel manufacture, railroad transportation, shipping, industrial construction, and other allied fields usually come from persons who were employed in those fields before going into military service. These individuals can obtain amazing amounts of accurate information both from what they see on the photographs and what they piece together from indications and discernible features associated with certain activities which may actually be hidden. Photo-interpretation provides a very accurate and extensive source of information relative to the strength, disposition, and activities of the enemy in areas not readily accessible to ground observers.

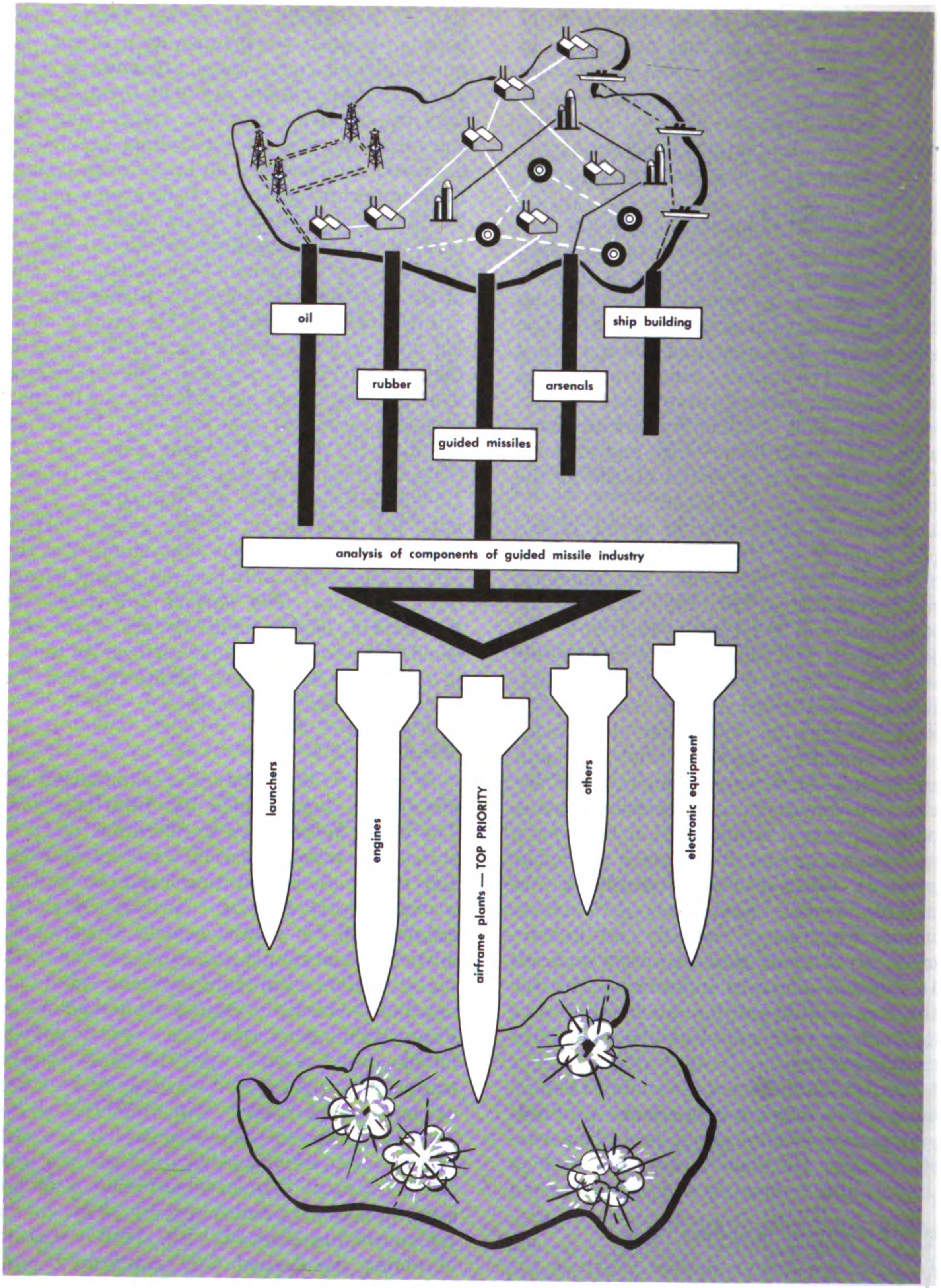
A detailed knowledge of the enemy's habits, customs, and military techniques also aids the interpreter in obtaining information from photo studies. He must know what to look for and be able to identify what he finds.

Priority Assignment to Surface Targets

To achieve a certain desired effect on the battlefield, some targets are attacked and others ignored. Because of a changing situation, a previously unmolested target may become of great importance as a target. In practically all cases, there are more targets available than missiles to fire at them. Therefore, in any given situation, targets must be considered in view of the factors previously discussed, and assigned an importance rating. Each important industry of the enemy nation is rated as a potential target system on the basis of the considerations mentioned above. The various component subsystems of the target systems receiving the higher-priority ratings are in turn rated by the same screening process. Top priority for missile attack is then assigned to one or more of the subsystems in an attempt to cripple or destroy the finished product output of that industry. Notice in the diagram on the following page that the higher ratings have been assigned to rubber, guided missile, and oil industries. Next, note the breakdown of the guided missile industry into the various subsystems. For the sake of simplicity, only one component breakdown is given. The various subsystems of the guided missile industry are then analyzed and assigned priority ratings. As shown in the diagram, airframe plants received the top priority for missile attack.

Post attack Target Analysis

It is important that an examination be made of a target which has been attacked. The target is examined to determine whether or not the missile mission was accomplished and to obtain information valuable to subsequent preattack target analysis. Such an examination for missile attacks is relatively difficult. Cameras placed in piloted reconnaissance bombers photograph bomb hits and much of the resulting damage from the attack. Photographic equipment in a one-shot guided missile would of course be useless. Television equipment might possibly give information up to time of impact but would be of no value in missiles whose ranges exceeded that of the television transmitters. In the employment of air-to-surface missiles, some information could be obtained by observers in the mother aircraft. But the best means of obtaining the desired data are from photographic reconnaissance at various intervals following attacks, prisoner-of-war interroga-



Target systems analysis and target selection

tion, captured documents, signal communications, and secret agents.

Knowledge of the physical damage to the various structures in a traget area gives information relative to particular warhead and fuze characteristics, method of attack on the target, and methods of target location, but such infomation is not sufficient, It is highly desirous to know the production loss and the effect produced on both the civilian and military personnel. The progress of reconstruction and repair must be followed to discover the enemy's order of repair priorities and to determine the time for reattack.

This post-attack analysis also provides the basis for future priority assignment to hostile targets; that is, it helps determine whether attacks should continue against aircraft factories, for instance, or whether they should be shifted to oil refineries or some other target system.

AERIAL TARGET CONSIDERATIONS

The remark is often made that "the best defense is a good offense." Likewise, the only completely successful defense against enemy missiles would be a vigorous offensive action that would prevent all enemy missile launchings through the destruction of unfired missiles, launching sites, supply lines supporting missile firing units, and missile producing facilities. However, prospects of such an offensive in the event of war are slight. Thus, consideration must be given to utilization of surface-to-air and air-to-air missiles as defensive weapons against attacking enemy missiles while in flight.

The anticipated high speeds, high operating altitudes, and evasive maneuverability of some aerial war weapons have surpassed the capabilities of present-day antiaircraft artillery as an effective defensive weapon. Design, construction, and employment of defensive missiles are therefore important aspects of the research and development program.

Areas or installations to be defended must be assigned defense priority ratings. Such a rating for a particular area is derived on the basis of estimates of enemy tactics to be employed against that area, importance attached to it by the enemy, vulnerability to enemy weapons, and recuperability factors. Some idea may then be formulated as to maximum allowable enemy effort and defensive measures necessary for effective protection.

The initial requirement of an air-defense unit is the volume of fire that the unit must be capable of delivering on the expected enemy effort. The volume of fire determination is dependent on characteristics of the proposed defensive missile and its kill probability relative to certain aerial targets. Consideration of estimated enemy tactics, area to be defended, and missile characteristics are aids in determining the damage probability that can be expected on an aggressor target. These factors also make possible a mathematical estimate of the number of missiles necessary to produce the desired target-destruction probability.

The various estimates and mathematical probabilities relative to anticipated enemy airborne attacks on friendly installations serve as a basis or guide in setting up guided missile defenses against attacking piloted aircraft and/or missiles.

Aerial Target Detection and Evaluation

The problems of utmost importance in a real-life situation are detection and evaluation of attacking missiles. Detection must be achieved at ranges great enough to permit necessary defensive measures. Before such measures can be taken, the attacking missile must be identified, and it must be evaluated as to speed, altitude, and course. This is done by means of radar. This data must be coordinated with the launching and guidance equipment of the defensive missile so that its launching time, launching attitude, flight path, and other factors are properly related for a successful mission. Each piloted aircraft or missile in a formation must be considered a separate and distinct target. It is probable, therefore, that surface-to-air missile fire will be directed from a centralized point to assure proper volume of fire and assure that various firing units will be directed to engage a particular target in a group.

In connection with air targets, the phrase "range of accomplishment" is frequently used. Range of accomplishment of mission refers to how near an attacking target must approach

a defense installation in order to succeed in its attack. Of two targets otherwise equally difficult to destroy, it may be highly advantageous to destroy one target at a greater range than the other. For instance, a piloted aircraft carrying guided missile to be launched at relatively long range has a longer range of accomplishment than a bomber carrying unguided gravity bombs. The craft carrying the guided missile should therefore be destroyed at longer range before release of the missile. When the missiles are released, each one will become a new target within itself — faster, smaller, and more difficult to destroy than the original craft.

Post Attack Target Analysis

Post attack analysis is also an important consideration where enemy airborne targets are concerned. If three firing crews defending a friendly installation are launching surface-to-air missiles at the maximum rate of fire, those missiles fired after the aggressor target is destroyed are useless, expensive expenditures. The difficulty arising from this problem varies with the use of surface-to-air and air-to-air missiles. In the case of air-to-air missiles, the range is usually much shorter, permitting visual or photographic observation of the engagement from the mother aircraft.

MISSILE SELECTION FOR A SPECIFIC TARGET

The selection of a particular missile for attacking some surface or air target actually begins with consideration of anticipated targets. In other words, target characteristics must play an important part in the design and construction of various missiles if proper performance is to be expected. This is especially true insofar as guidance is concerned. Some of the criteria that are evaluated as aids to missile selection are discussed here in a general manner rather than going into various specific missiles and the targets for which they were designed.

Target Considerations for Missile Selection

Some of the target considerations on which missile selection is dependent are speed, maneuverability, range, vulnerability, and identification.

SPEED AND MANEUVERABILITY. Speed and maneuverability of any enemy target which we may attempt to destroy are two factors which likely will always be fairly well known to us. Any new missiles, vessels, or vehicles developed by a foreign country could not conceivably have a radical change in these two factors from those already known to exist; that is, any future change in maneuverability and speed will not be to the extent that we cannot quickly and easily determine these two factors and then compensate for them in our own defensive or offensive equipment. Of course, the sudden appearance of a target with the speed and maneuverability of the so-called "flying saucers" would be quite another thing, but we can only deal with what we know to be true.

Targets can be grouped in accordance with speed and maneuverability, as follows:

Group I Targets. This group consists of area-land targets having zero speed and zero maneuverability. An example of such targets is large cities which are thickly populated and highly industrial. Due to the large size and the fact that the target is not moving, guidance information can be preset into the missile before launching with reasonable certainty of hitting within a certain area. Such targets make the least demands on a missile for accuracy. Of course, the higher the speed of a missile selected for use against area targets, the greater is the possibility of the missile eluding enemy countermeasures. The German V-1 and V-2 used against London are good examples of the value of speed. The slow speed of the V-1 permitted effective countermeasures. This was not true in the case of the faster V-2, once it was successfully launched.

Group II Targets. Placed in group II are point-land targets, ships, submarines, trains, and other moving targets. The speed range of targets in this group goes from zero in the case of point-land targets to approximately 100 miles per hour in the case of some moving targets. Even though the maneuverability of some of the targets may be relatively great, the comparatively slow speed would still allow a direct hit or near miss if attacked by the right type of missile. Missiles used to attack group II targets need a much more accurate guidance

system than missiles used for group I targets. The missile used should be capable of being launched from an aircraft or from a surface on a predetermined trajectory. It should also be capable of being guided in flight by means of radio signals, radar beams, or homing head to compensate for errors or for movement of the target. An example of a missile selected to engage a maneuvering ship is an air-to-surface missile with a radar homing guidance system.

Group III Targets. Propeller driven aircraft and possibly certain guided missiles, which are limited to subsonic speeds are examples of targets in group III. A missile selected for destruction of such targets must have greater speed and possess more sensitive and accurate guidance equipment than missile selected to attack targets in groups I and II. The guidance system may be basically one of the types listed under group II but must compensate for the added speed and maneuverability of targets found in group III.

Group IV Targets. Generally speaking, jet-propelled aircraft and supersonic guided missiles are placed in group IV. Missiles selected to engage group IV targets must be capable of supersonic speeds, high altitudes, and guidance-system sensitivity and accuracy to meet the given situation.

RANGE. Range is another consideration in missile selection. Range is basically the distance between the launcher and the target. In the case of surface targets this range is more or less fixed. If, however, a particular SSM satisfies all other requirements for attacking a surface target except the range, the launcher must be relocated at a point closer to the target under consideration. Utilization of the air launching technique is another means of increasing the range of certain guided missiles. In any case, adequate range is a necessity.

VULNERABILITY. Considerations as to the vulnerability of targets were discussed earlier in this chapter. In selection of a missile to attack various targets, these considerations must be weighed against the efficiency of the warhead and the arming and fuzing system of the missile. Consideration must also be given to the number of missiles of a particular type necessary to ensure destruction of a given target. One particular target might be highly vulnerable to blast-type warheads while another would be more vulnerable to a fragmentation warhead. Vulnerability can now be defined as the probability of effective damage being caused to a target by a given missile at a given burst location with respect to the target.

IDENTIFYING CHARACTERISTICS OF TARGET COMPONENTS. Certain characteristics of the components that make up a target are valuable aids in missile selection from the standpoint of guidance. For example, if a target component emits some form of radiant energy, a missile utilizing a homing guidance system sensitive to the particular form of radiant energy may prove more effective than a missile guided by some other method.

Other Factors Which Influence Missile Selection

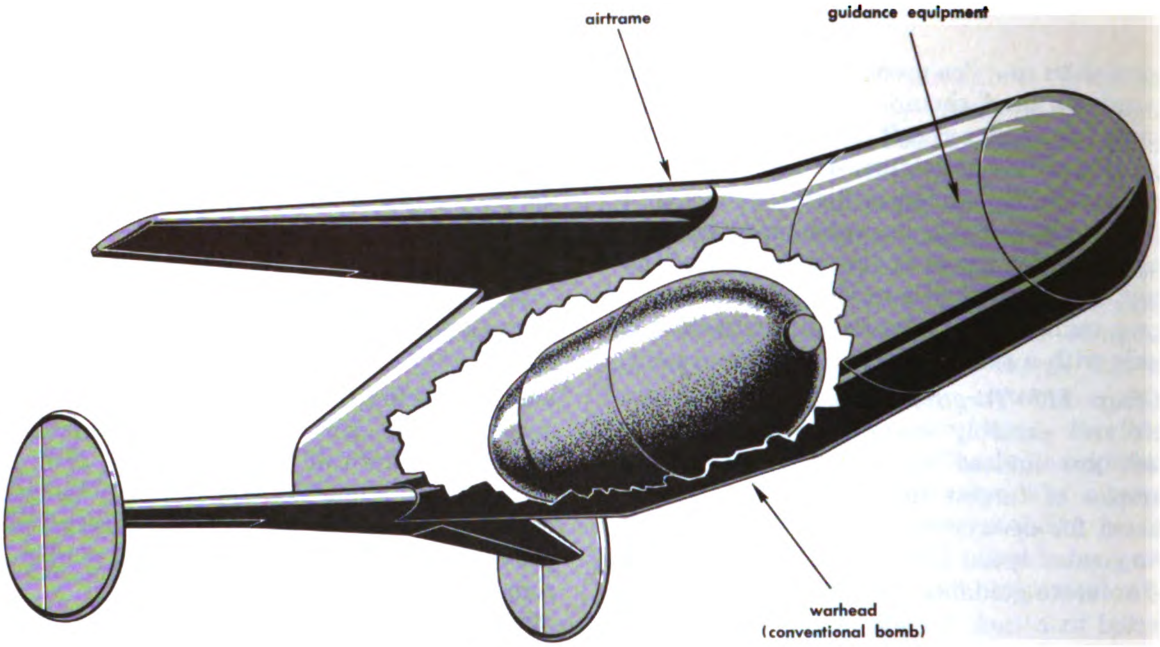
Another factor that must be considered is rate of fire based on the availability of a particular missile and handling and launching procedures. Still another factor is dependability of the missile under varying temperature, weather, and field conditions.

After consideration of all influencing factors, a missile is selected, or designed if none exists, that most nearly meets the qualifications necessary to achieve the desired result against a target with a given set of characteristics.

MISSILE WARHEADS

The object of the design and construction of any missile is to deliver a warhead to a designated target. The missile is just the transporting vehicle. However, the most accurate guidance and propulsion systems are of limited value if the warhead cannot produce sufficient lethal effect at the right time to destroy or incapacitate the target.

During World War II guided missiles, for the most part, adopted standard bombs as their explosive payloads. These bombs, encased in suitable airframes, often equipped with some form of guidance system, constituted our first missile. The sketch on next page is an example of such an arrangement. It shows a standard bomb utilized as a warhead, an airframe for transport, and radar for guidance.



Employment of a conventional bomb as a warhead in an early missile

Constant research and experimentation in warhead design make it possible to ascertain optimum effect of warheads on targets of various natures. The warhead selected for a given target depends on the characteristics of the target. Kinds of armor found on the target, penetration desired, speed of target. type of destruction desired, and other factors must be considered in warhead design. Special design considerations apply to each type of target, but surface-based targets offer the fewest differences from conventional weapons. It is the aerial target which imposes maximum difficulties upon warhead design and missile components.

Desirable Characteristics for Warheads

Just as in the case of any other missile component, there are certain characteristics desirable in any warhead. However, all warheads do not possess all the desired properties listed below:

a. Warhead design should permit interchangeability with other type warheads and rapid installation onto the missile under field conditions.

b. The warhead components should be capable of being readily assembled under field conditions.

c. The warhead should be able to withstand accelerations of the missile used, and the filling agent should not deteriorate during flight.

d. The warhead design should be such as to obtain maximum distribution of destructive elements on the target area.

Types of Warheads

The word *warhead* usually implies that a quantity of high explosive is encased in a metallic shell. This is not always true. Depending on the target, a warhead may contain high explosives, incendiaries, bacteria, or gas. Also, in the high-explosive-warhead category, the warhead can be designed to produce varied results. It may be designed to produce destruction by concussion or by flying fragments of the warhead casing. The following discussion points out the various types of warheads from the standpoint of basic physical construction and effects produced on various targets.

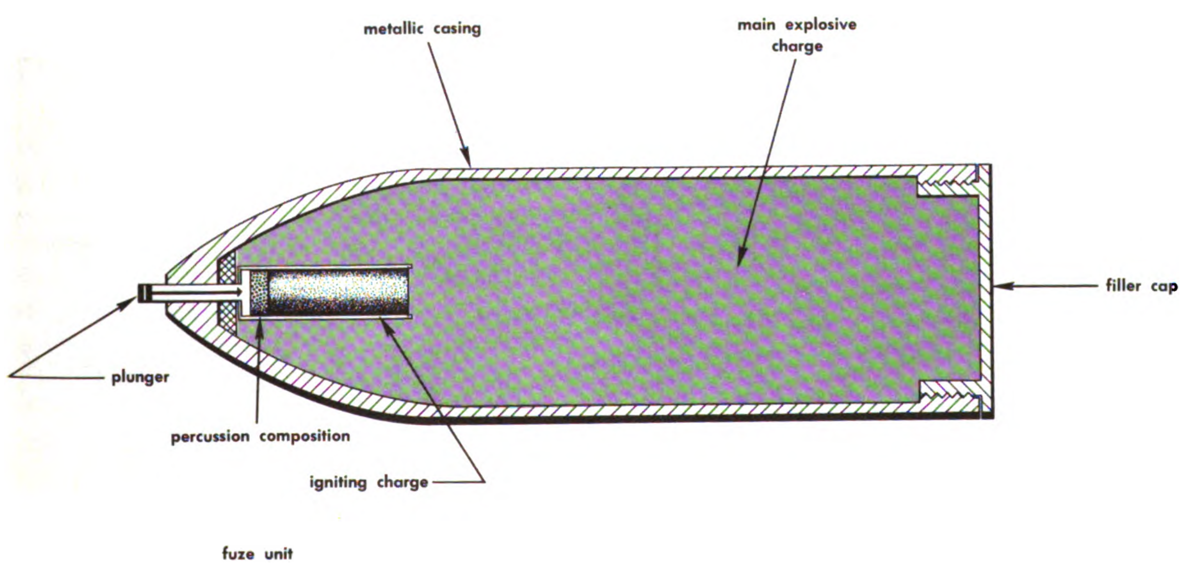
BLAST-EFFECT WARHEAD. The blast-effect warhead consists of a quantity of high-explosive material housed in a metallic case. The force of the explosion sets up a pressure wave in the air, or other surrounding medium, which causes damage to the target. This action is somewhat analogous to the waves that result when a large stone is dropped in a pool of water. The blast-effect warhead is

most effective against underwater targets because of the greater density of water as compared with the density of air. However, this type has been successfully used against ground targets. It is even less effective against aerial targets than ground targets, because of the lessening of air density with altitude. The following illustration shows a metallic casing, containing explosive materials and a suitable fuze unit for initiating the major action within the warhead. Various fuzes are discussed later in this chapter. On impact with the target, the plunger strikes and detonates the percussion composition which fires the igniting charge. The igniting charge causes the main explosive charge to detonate and produce destructive compression waves in the surrounding medium.

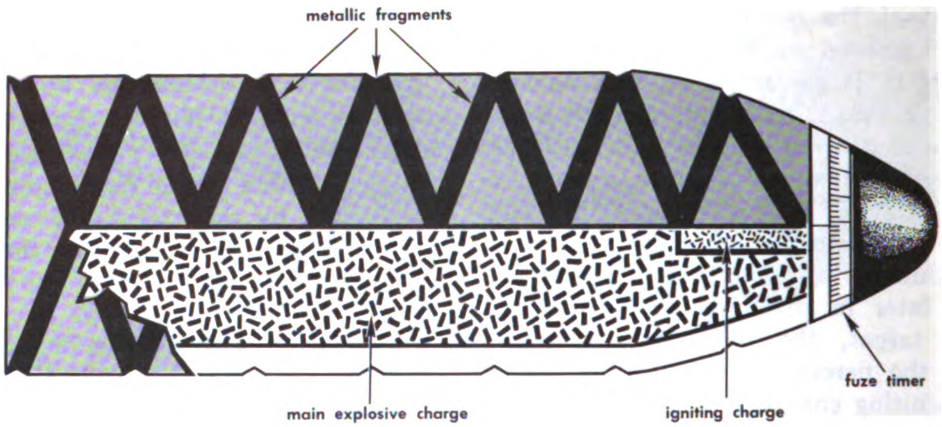
FRAGMENTATION WARHEAD. A fragmentation warhead utilizes the force of an explosive charge to eject metallic fragments of the casing at high velocity to achieve target damage. The size of fragments, velocity of fragments, and spray pattern formed by the flying fragments can be controlled by variations in design and construction of the warhead. The size of fragments is controlled by weakening the metal casing at selected points, causing fractures to occur at these points first when the warhead explodes. The velocity of the fragments is controlled by the type of explosive used and by the explosive-to-metal ratio. The spray pattern formed by the fragments is determined by the point of detonation and shape relationships of the warhead.

Performance of the fragmentation warhead is limited by the amount of metal available to form fragments and by the amount of explosive that is available for breaking the metal into fragments and for giving them the necessary velocity. The illustration on next page shows a metallic shell designed to form fragments of the desired size. It is loaded with an explosive charge and contains a fuze unit that causes detonation at a predetermined distance from the target.

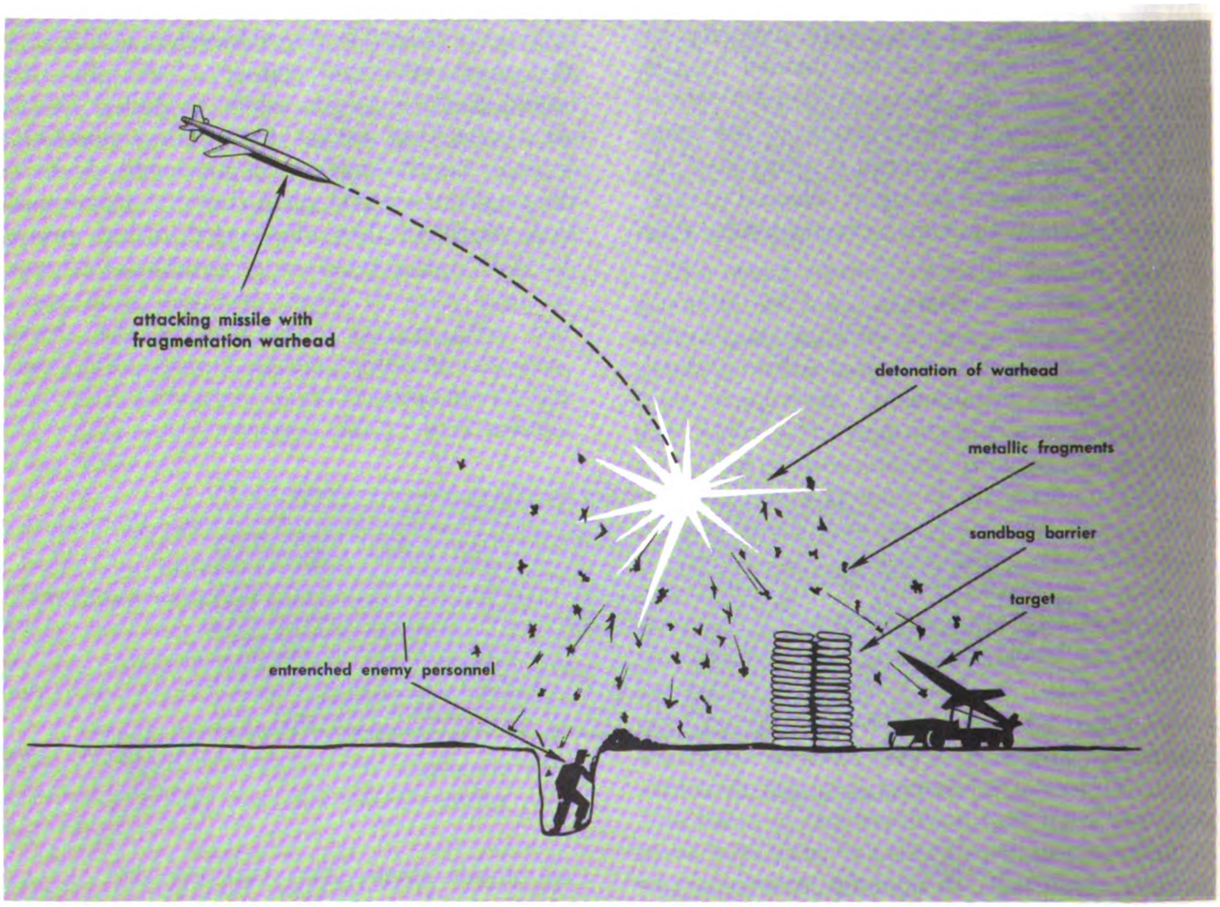
Aerial targets are more susceptible to damage by fragments if the warhead explodes a reasonable distance from the targets than if a similar warhead explodes upon contact with them. A fragmentation warhead is also more effective against a partially protected surface target when exploded in the air above that target than when caused to explode on the surface of the earth. Note in the diagram on next page that the fragmentation warhead has detonated some prede-



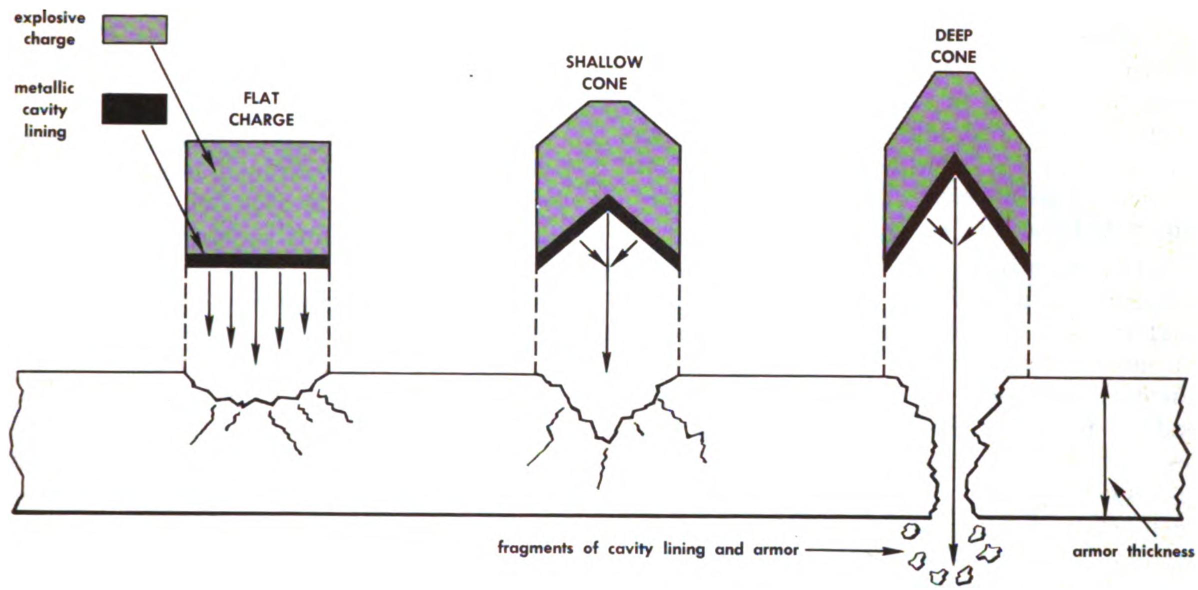
Basic construction of a blast-effect warhead



Basic construction of a fragmentation warhead



Effect of fragmentation warhead on surface targets



Effects of various charge shapes

termined distance above the targets. Fragments resulting from the explosion strike both the partially protected missile and the entrenched infantryman.

SHAPED-CHARGE WARHEAD. A shaped-charge consists of a casing containing highly explosive materials and is so designed that the force of the explosion is pointed, or concentrated, in one direction. Damage is produced by the compression force of the blast and/or by fragments ejected in the direction of the blast. The shaped charge is most widely used against armored surface targets. For example, the antitank bazooka rocket used during World War II and in Korea utilized the shaped-charge design. The figures above illustrate the influence that the shape of a charge has on the distribution of available explosive force. As you study the figures, assume that the quantity of explosive material is the same in all three cases. Only the configuration of the explosive material is shown, not the entire warhead construction. The arrows indicate the direction and magnitude of the explosive force.

Note that the *flat charge* to the left of the three figures shows the explosive forces distributed rather evenly over a given area of the target, with no appreciable amount of penetration. The *shallow-cone* middle figure shows greater concentration of the explosive force in one direction and deeper penetration of the target. The *deep-cone* charge to the right has produced adequate concentration of explosive forces for complete penetration of the armored target. Metallic fragments of the warhead and armor can now reach the interior of the target to do additional damage.

Use of the shaped-charge warhead against aerial targets is also under investigation. The warhead could be gimbaled and constantly positioned relative to the target by means of electronic equipment. If detonation time is properly controlled, ejected fragments would be concentrated on a collision course with the target.

EXPLOSIVE-PELLET WARHEAD. An explosive-pellet warhead consists of a group of separately fuzed explosive pellets housed in a casing. The casing contains an additional quantity of explosive material, or other media, which supplies the force necessary to eject the explosive pellets from the main warhead casing. The pellets do not explode until they contact or penetrate the target. In the case of aerial targets, maximum destruction is accomplished when the pellets are detonated after penetrating the target. Each pellet contributes both blast effect and high-velocity metallic fragments when detonation occurs.

In the diagram below, explosive pellets have been ejected from the warhead of the defensive missile. Note that after ejection from the warhead in the nose of the missile, the pellets are still intact as they approach the target missile. After penetrating the target missile, the pellets explode, causing both blast and fragmentation damage.

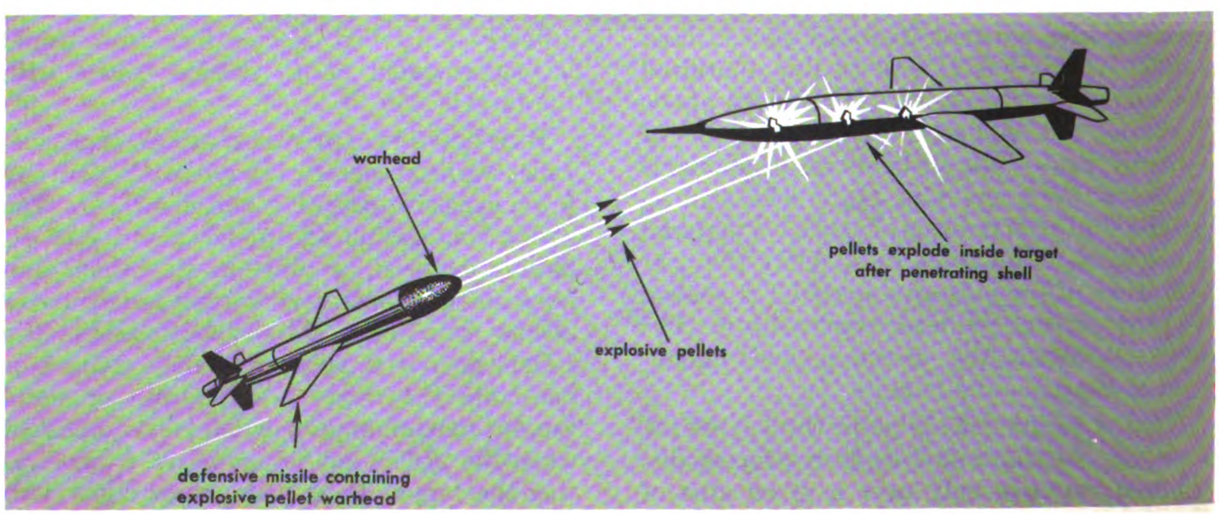
CHEMICAL WARHEADS. A chemical warhead is designed to produce personnel casualties by inhalation of, or by physical contact with, poisonous materials, or it is designed to destroy combustible targets by use of incendiary materials. Previously, poisonous war gases have been classified according to tactical use, physiological effect, and persistency. More specifically, war gases are referred to as casualty gases, vomiting gases, nonpersistent gases, and persistent gases. The type and amount of gas placed in a missile warhead depend on the characteristics of the target and the results desired.

Incendiary warheads contain chemicals that combust violently, produce extremely high temperatures, cover a large area when released, and are difficult to extinguish. Incendiary warheads are used mainly against surface targets, but also are effective against aerial targets containing inflammable materials. Magnesium, which burns at a temperature of approximately 2000° C is a good example of an incendiary. Other substances that may be used are jellied gasoline, jellied oil, and phosphorus, The type of incendiary material used in a warhead and the point of warhead detonation depend on the nature of the target.

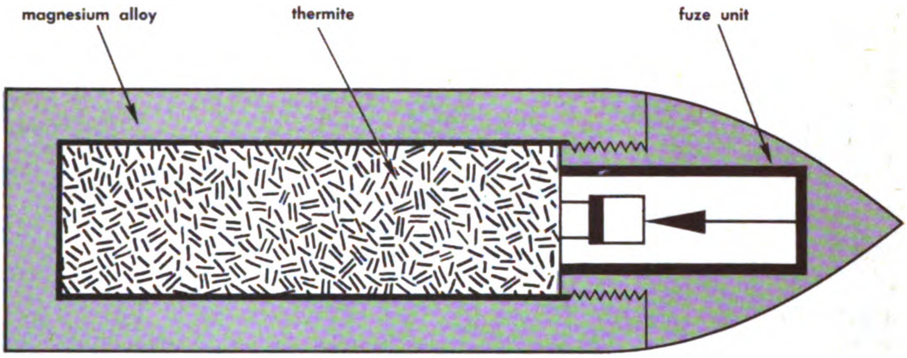
The basic construction employed in a magnesium incendiary warhead is illustrated on the right. It consists of a hollow magnesium alloy casing packed with thermite (aluminum and iron oxide) and a fuze with a suitable igniting mixture (powdered magnesium or black powder). The fuze ignites the thermite which burns violently. The burning thermite ignites the magnesium casing itself. An explosive charge may also be placed in the warhead to produce greater dispersion of the incendiary materials over the target area.

BIOLOGICAL WARHEADS. A biological warhead contains living organisms, or agents, that cause sickness and death. An explosive charge placed in a biological warhead would insure ejection and initial dispersion of the biological agents. Special attention must be given to design and construction of biological warheads in order that the bacteria will remain alive and be carried to the target under the most favorable conditions.

ATOMIC WARHEADS. For reasons of national security, the amount of material available for publication relative to atomic warheads is necessarily limited. Thus, no attempt is made here to present information pertaining to the physical structure or assembly of components of an atomic warhead. Instead, the general



Separately fused exploding pellets after penetrating target



Basic construction of a magnesium incendiary warhead

characteristics and effects of an atomic explosion are discussed.

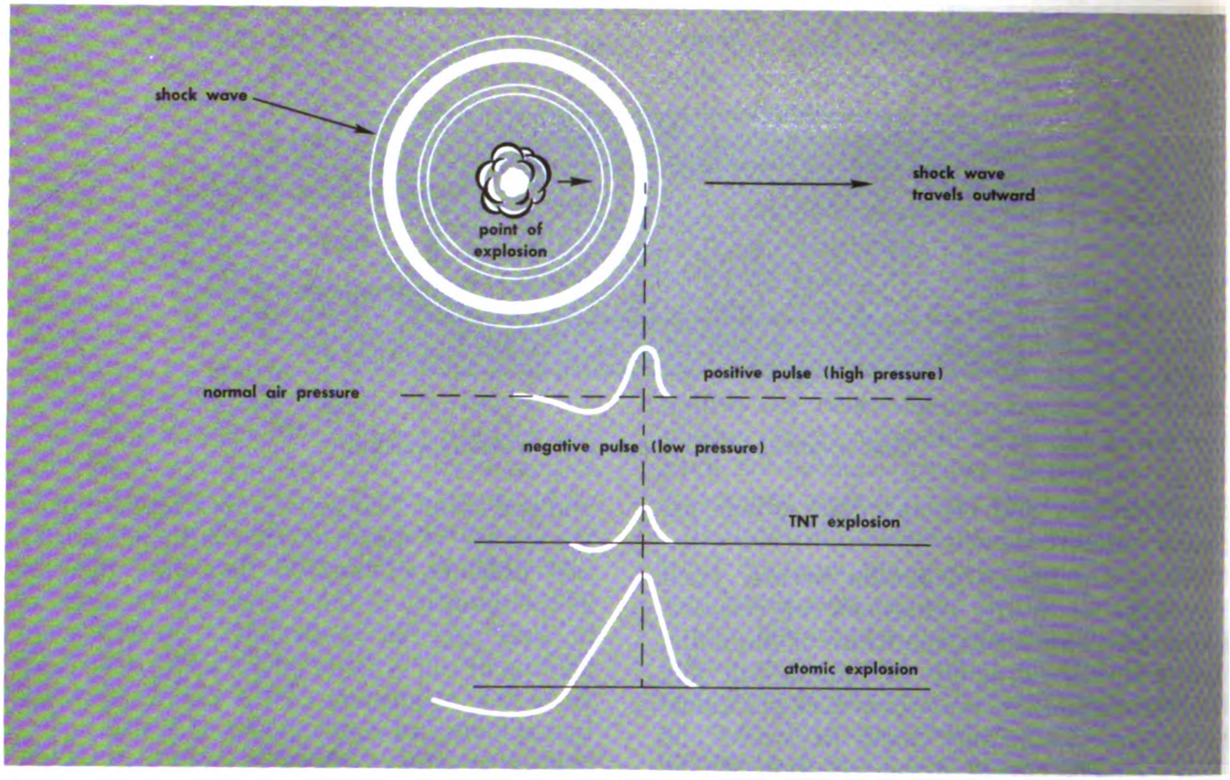
When an atomic explosion occurs, great quantities of deadly radiations are released. Exposure to sufficient quantities of the radiation will kill as surely as bullets will kill. Exposure to atomic radiations may kill quickly, or the victim may linger for weeks and then die. The effects of the radiation depend on the dosage. A low concentration for a long period may be just as fatal as a high concentration for a much shorter time.

A great amount of heat is developed at the instant of an atomic explosion. In the first few millionths of a second, after the explosion is initiated, a great "ball of fire" develops. The temperature of the center of the mass of hot gases reaches millions of degrees centigrade. The fire ball dissipates by rising vertically and radiating its energy just as the sun radiates energy. The energy is radiated as visible ultraviolet and infrared light. The visible light can cause temporary blindness. The ultraviolet light can cause skin injuries of the same type as those of sunburn, although much more severe. The infrared can kill by producing flash burns. At the same time, the heat causes combustion in easily inflammable materials up to a mile or more from the blast, depending on the size of the nuclear weapon.

Because of the temperature and pressure of the gases generated by an atomic explosion, a destructive shock wave develops. The shock wave consists of two pressure regions, one higher than normal atmospheric pressure and one lower. The wave moves outward from the center of the explosion with a velocity and pressure differential that vary directly as the intensity of the explosion. The diagram on next page shows the formation and propagation of the types of shock waves described. Note the magnitude of the shock produced by exploding an equal weight of TNT.

Let's consider briefly what happens when an atomic bomb is detonated. The atomic explosion illustrated in the diagram on page 481 is initiated when a neutron (electrically neutral particle of matter) is caused to enter the nucleus of a uranium atom. The uranium atom "blows-up," ejecting two lighter atoms and a number of particles of matter including two or three additional neutrons. Nuclear fission is the term used to denote this atom-splitting process. About 200 million electron volts of energy are given up for each atom which undergoes fission. Since the fission reaction is produced by a neutron, and since the reaction throws out more neutrons, other uranium atoms of the original mass are split, producing two or three more neutrons each. Thus the reaction grows to explosive proportions within a very small fraction of a second.

The effects of an atomic explosion are divided into two phases, as indicated in the diagram on page 481. Phase I consists of those phenomena which accompany the actual explosion. Pressure (shock wave) has been discussed in detail. Alpha and beta particles are also emitted at the instant of the explosion. While they are harmful, their range is so short that they need not be considered a



Comparison of shock waves formed by equal weights of TNT and fissionable material

serious threat. Many of the electromagnetic radiations such as infrared visible light and ultraviolet radiations have been discussed previously. The danger of X-rays is well known from the standpoint of body penetration. Gamma radiation has the same general characteristics as does X-ray radiation, but it is more energetic and will penetrate greater thicknesses of shielding materials.

Phase 2 consists of radiations which persist after the explosion. Let's consider these radiations. Unfissioned radioactive material (uranium 235) is reduced to a powder and scattered by the explosion. The alpha particles emitted by this material prove to be a great hazard if taken into the body through the mouth or through open cuts.

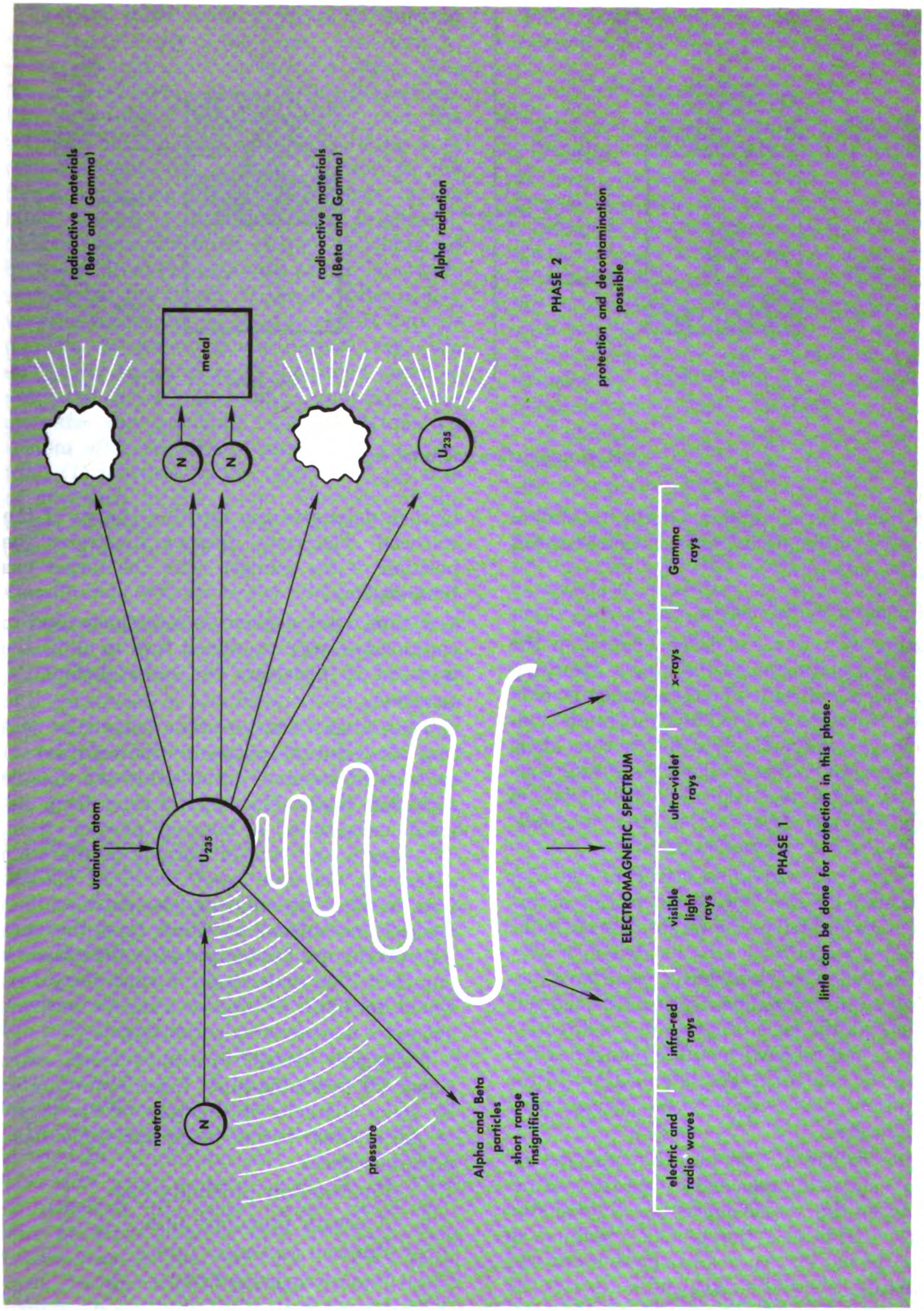
That portion of the original matter which undergoes fission produces radioactive materials that are scattered by the explosion. The particles emit beta and gamma radiation. As previously stated, the beta particles are not particularly dangerous. On the other hand, gamma radiation may be dangerous over a large area after the explosion.

A hazard due to radioactivity induced in the components of the target area also develops from an atomic explosion. Induced radioactivity is produced by the neutrons which are released at the instant of the explosion. These neutrons enter the nuclei of some atoms found in some target components, causing a physical reaction which leaves a radioactive atom as a product. Considerable gamma radiation has been known to result from this action.

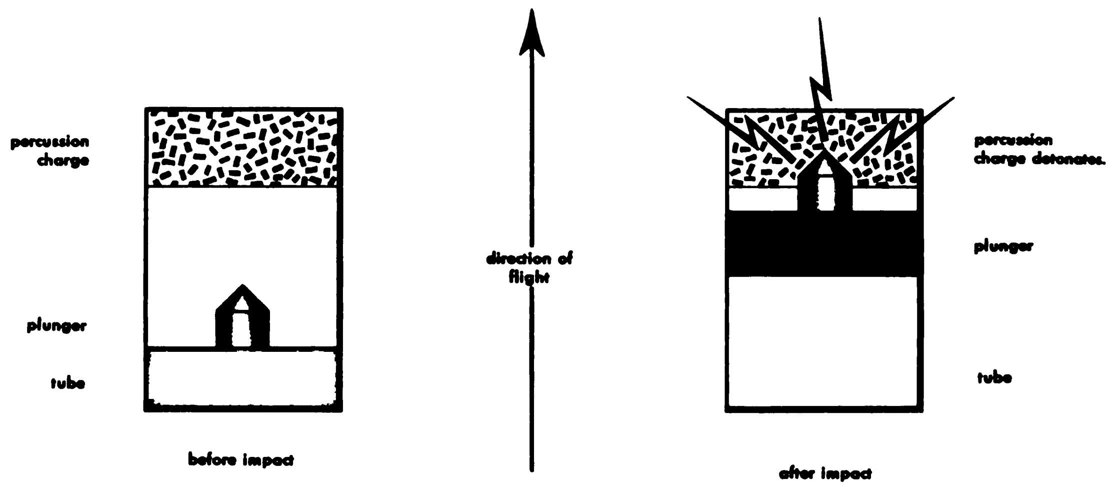
Atomic warheads complete the line-up of warheads used in missiles. The hydrogen bomb will add still another kind of warhead to the types of warheads discussed here.

FUZES FOR MISSILE WARHEADS

A fuze is a device designed to activate the bursting charge in a given missile warhead at the proper time for maximum effectiveness against a particular target. Special design considerations apply to each type of target. However, surface-based targets necessitate the least differences from current practice with fuzes on more conventional weapons. On the



Activity stemming from atomic explosion



Impact fuse before and after impact

other hand, aerial targets impose maximum difficulties on the design of fuzes and warhead components.

Description of Various Fuzes

Fuzing of a warhead is dependent on the overall characteristics of the target, attacking missile, and warhead. The target must be considered from the standpoint of location, vulnerability, speed, and physical structure, along with the capabilities of the missile and warhead, in order to make the probability of damage to the target as high as possible. Various fuzes are discussed below.

IMPACT FUZE. An impact fuze is one that is actuated by the inertial force that occurs when a missile strikes the target. For example, consider a cylindrical tube located in a warhead with a shock-sensitive explosive-percussion charge permanently fixed in the forward end of the tube (relative to the direction of flight) and a heavy metallic plunger at the rear end of the tube. The left diagram in the illustration above shows the arrangement just described.

While in flight, the movable plunger remains against the rear end of the tube. However, when the missile comes to a sudden stop against the target, the plunger, tending to remain in motion, rushes to the forward end of the tube where it strikes and detonates the shock-sensitive fuze charge. The fuze charge in turn detonates the bursting charge of the warhead. The right diagram illustrates the position of the plunger after the target has been engaged.

A time delay element is sometimes used in conjunction with an impact fuze to allow the warhead to penetrate the target before detonation.

TIME DELAY FUZE. A time delay fuze is a fuze designed to detonate the warhead at some predetermined time after the missile flight has begun. One type of time delay element consists of a burning powder train. The length and rate of burning of the train determine the time before warhead detonation. Another type of delay element resembles a watch mechanism. In either type, the time interval can not be changed after the missile is launched. For that reason, a time delay fuze would not likely be used on a missile warhead against a maneuvering aerial target.

PROXIMITY FUZES. Proximity fuzes, often called VT (variable-time) fuzes, are actuated by some characteristic feature of the target or target area. Listed below are some basic proximity fuzes:

1. Photoelectric proximity fuze.
2. Accoustic proximity fuze.
3. Pressure proximity fuze.
4. Radio proximity fuze.
5. Electrostatic proximity fuze.

Each of these fuzes is preset to function when the intensity of the target characteristic, or target-area characteristic, to which that fuze is sensitive, reaches a certain magnitude. Proximity fuzes are designed so that the warhead burst pattern will occur at the most effective time and location relative to the target. Adapting the proximity fuze to the burst pattern of the warhead is, in general, difficult since the burst pattern is influenced by the relative velocity with which the missile approaches the target. If targets with widely varying speeds are to be attacked, it may be possible to automatically adjust the fuze sensitivity on the basis of the target speed as predicted by a computer. Proximity fuzes activate the auxiliary warhead-detonating system after electrically integrating two factors: (1) nearness to the target, and (2) rate of approach to the target. However, if the target succeeds in jamming a proximity fuze, only a direct hit can be effective.

Classification of Fuzes by Position in Warhead

When classified according to assembled position in the warhead, fuzes are either *point fuzes*, which are assembled in the nose of the warhead, or *base fuzes*, which are assembled in the rear of the warhead. As previously stated, the fuze or combination of fuzes used in a warhead and fuze location depend on the mission at hand and the effect desired.

Warhead Detonation Points Relative to Fuze Type

Now that we have discussed the various types of fuzes from the standpoint of definition, structure, and operation, let's look at the influence of various fuzes on the point of warhead detonation. Notice the drawing on the following page.

In the case of an impact-fuzed warhead (top drawing), the explosion occurs when the missile strikes the target.

In the middle drawing, a time delay element is used in conjunction with an impact fuze. The time delay element allows the warhead to penetrate the target before detonation. The principle involved in a time-delay fuze is the same as that involved in the impact fuze. In the third drawing, the proximity fuze has been actuated by some characteristic of the target, causing warhead detonation at a predetermined distance from the target.

MISSILE LAUNCHING SITES

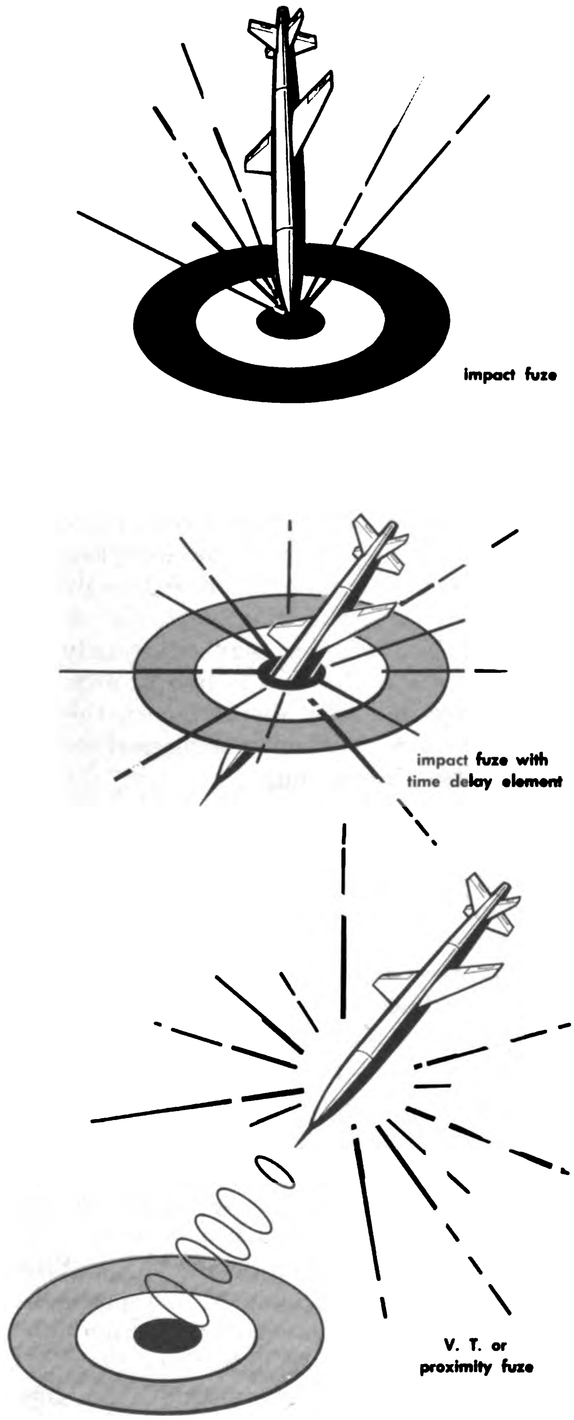
A launching site is the area maintained by a firing unit from which the missiles are placed in final readiness and fired at designated targets. The sites should be well dispersed as a defensive measure against enemy bombardment. Dispersed sites also permit missiles to be launched simultaneously from different launching sites without one launching interfering with another. The launching site should be easily accessible by railway and/or highway vehicles to allow for rapid transportation of missiles, propellants, and related equipment.

In the selection of launching sites for missiles using boosters, care must be exercised to prevent booster assemblies from falling among friendly troops. Launching sites for short-range offensive missiles may necessarily be located very near the battle line in order to exploit range, but they should be arranged in depth to provide continuity of support should forward sites be overrun.

In any case, launching sites will be considered prize targets by the enemy. For this reason, the more extensive the use of camouflage, and the speedier and more often changes in location can be made, the more difficult it will be for the enemy to identify and eliminate these sites. However, rapid and frequent relocation of a launching site is limited by the size and complexity of the launching equipment. Nevertheless, alternate locations for launching sites should be under consideration at all times.

Launching-site selection and offensive and defensive utilization of that site vary with the type and properties of the missile for which the site is designed. Some of the problems that arise in connection with launching sites for surface-to-air missiles differ from these encountered with surface-to-surface missile launching sites. Also, launching of short-range missiles near enemy held territory presents problems different from those that arise when launching long-range missiles from sites located on the home front.

You can readily see that this discussion has attempted to present only the general



Influence of fuze on point of warhead detonation

considerations given to launching sites rather than specific data on individual missile launching sites.

MOBILE LAUNCHERS

Mobility in launchers is a highly desirable characteristic. An ideal mobile launcher is one that can transport one or more complete guided missiles, allied equipment, and necessary personnel to a predetermined site and then ready, aim, fire, and, if necessary, guide that missile to a selected target.

You have no doubt formulated a mental picture of mobile launchers equipped for land and air travel. Of great importance are launchers equipped for the sea. The potentialities of mobile launching sites at sea are tremendous. Any area on earth with potential target characteristics is within approximately 1700 nautical miles of open-sea areas available for practical naval operations. Large ships and submarines for launching missiles are certain to fit into future joint employment of missiles.

SUPPORT FACILITIES

Now that we have discussed some of the aspects of target selection, missile selection, payload, and launching sites, let's briefly consider some of the facilities that must be maintained to support a missile launching.

Combat Squadron Organization

A missile squadron must be organized around each specific type of missile, after that particular missile has been accepted for use by the United States Air Force. For this reason, a missile squadron organized around a particular type of missile will differ in the finer details from a squadron organized for a missile of another type. However, certain basic organizational principles hold true for almost all missile squadrons. The following discussion covers squadron organization in general with no detailed information on any particular missile or number of personnel. The missile squadron is discussed in terms of the following three basic units: technical unit, service unit, and firing unit: The technical unit and service unit for one type of missile squadron function approximately the same as they would for any other type of missile squadron, but this does not hold true for the firing unit. The firing unit for air-launched missiles, for example, employs equipment and methods of operation very different

from the firing unit for ground-launched missiles.

MISSILE TECHNICAL UNIT. A technical unit of a missile squadron is the receiving center for packaged missiles, or sections, component parts, and allied equipment, from the producing agencies in the zone of interior. Major component parts of missiles and allied equipment are assembled, tested, and packaged at the missile factory to insure a high degree of reliability, after which they are shipped to the assembly area in the theater of operations. The technical unit must maintain an assembly and checkout area, ordnance area, liquid-propellent storage area, repair depot, and storage and supply depots for missiles and spare components. The various areas and depots should be located conveniently near a suitable seaport if the missile is being employed in an overseas theater.

At the assembly area, trained personnel unpack, inspect, assemble, and checkout various missile components and systems, and they prepare the missile for transport to the firing site. Checks are performed on guidance and control systems components, and any components or airframe sections damaged in shipment or faulty for other reasons are replaced. Faulty components are sent to the proper repair depot. Liquid-propulsion systems are run briefly to check for proper operation and system leaks.

In general, assembly at the assembly area includes everything except final fueling and pressurization of the missile, warhead and fuze installation, and final installation of certain guidance and control power supplies. The extent of assembly, of course, depends on the missile being used. For example, a certain solid-rocket-powered missile may be completely assembled and stored, or sent to the firing site, with warhead installation being the only operation necessary before firing. On the other hand, a missile which uses a liquid-propellant propulsion system may not be fueled until placed on the launcher at the firing site. This procedure would be followed for missiles using liquid oxygen. The high rate of evaporation of the substance would result in a great loss during transportation from assembly area to launching site. The same line of reasoning holds true for highly sensitive pieces of guidance equipment that may become damaged during transit.

A number of complete assembly lines must operate within the assembly in order to maintain the desired flow of missiles to the various launching sites. When assembly is completed, the missile is either placed in storage until needed or loaded on suitable mobile trailers and sent to the launching areas.

Upon arrival at the assembly area, the warheads, fuzes, and solid-rocket power plants are probably sent to an ordnance area. Here, trained crews assemble, check, and maintain storage facilities for the warhead and fuze units until transported to the firing site for installation in missiles. Other trained crews are responsible for checking, assembling, and storing solid-rocket motor units until delivery to the firing site. In some cases, it is necessary to maintain a solid-rocket propellant supply depot separate from the ordnance area, depending on the number, type, and rate at which solid rocket units must be handled. In all cases where the missile employed by the squadron uses liquid propellants, a liquid-propellant storage area must be separately maintained by properly trained personnel.

Even though the duties of the technical unit are related to both ordnance and supply, its job is primarily a technical one. The size of the technical unit depends on the state of assembly of the missile when received and on the complexities of the technical checkouts. The size and dispersion of the various areas and depots of the technical unit depend on the number, size, and type of the missile components and on the nature of propellants and warheads used.

MISSILE SERVICE UNIT. A service unit is responsible for supplying the firing units with propellants, warheads, fuzes, and allied equipment, as well as all nontechnical items. The service unit must obtain the missile and other equipment from the assembly area and various supply depots and transport these items directly to the firing site, or maintain a distribution point from which the firing units may draw.

This unit is mainly responsible for transporting missiles, propellants, and related equip-

ment. A secondary duty of the service unit is that of motor-vehicle maintenance. Missile and fuel trailers, and the motor vehicles used to haul these trailers, must be kept in operating condition.

MISSILE FIRING UNIT. The mission of a firing unit is to set up and maintain launching sites and to launch and control (if necessary) the flight of the missile. Launching-site personnel operate as teams, each especially trained to perform certain operations in preparation for launching. Specialists fuel and pressurize the propulsion system, make line pressure checks, and install booster units. Control and guidance teams make final *go*, *no-go* checks on the respective system components, and they make any necessary adjustments or settings in the equipment. An armament team installs the warhead and fuzing unit.

During the performance of these and other prelaunch operations, any malfunction detected is isolated, and the malfunctioning unit or component is replaced with a spare. The defective part is then returned to the proper repair depot. In general, no trouble shooting or detailed repairs are done at the launching site.

The design and properties of the missile determine whether or not it is placed on the launcher before, during, or after the various prelaunch operations. When all is in readiness, the missile is fired by remote control from a safe distance and guided to a predetermined target.

Other Supporting Agencies

The intent of the preceding discussion of support facilities was to describe briefly the various agencies responsible for, and the disposition of, a guided missile from factory to launching site. A number of other duties must be performed in support of those already discussed. For example, a headquarters section performs staff, administrative, and personnel-procurement duties. Security troops must guard the various installations against surprise attack or observation from land and air. Medical personnel familiar with injuries related to the handling of missiles, propellants, and explosives are necessary in any missile squadron. Still another important supporting facility is the network of communications that must be established and maintained between all the installations of a missile squadron. Needless to say, the mess personnel, generally thought of as the "major supporting facility" of any unit, are no exception in the case of a missile squadron.

MISSILE TACTICS TODAY AND TOMORROW

While certain missile tactics are certain to change slightly in the future because of changing conditions, missile tactics as presented in this chapter will continue to follow the present-day pattern. Improved techniques and improved equipment of the future will not alter the fundamental practices as you have learned them here.

Chapter 11

Guided Missile Instrumentation

Instrumentation as it applies to the guided missile field is discussed in this chapter. Instrumentation is thought of here as primarily involving systems for collecting structural and functional data from experimental missiles in flight. This data is collected to facilitate the research and development phase of the missile program.

The data obtained under actual flight conditions, when analyzed, is invaluable for the correction of structural or aerodynamic faults. Thus, perfection of missile design and operation is achieved in the shortest possible time and with a minimum of experimentation.

With these objectives, many systems of instrumentation have been developed or adopted for use in connection with missile research and development. Some of these systems, such as optical observation and radar tracking systems, are adaptations of those used for determining flight characteristics of conventional aircraft or for range finding and gun direction. Other systems have been developed specifically for *telemetering* missile data. More than one system may be employed within a missile, depending upon the nature and extent of the information required.

Regardless of the type of system employed, it is essential that the data obtained be as accurate as possible. It must be in such form that it can be readily decoded and permanently recorded for subsequent analysis and for comparison with data collected by other means or predetermined by the design engineers.

During the various stages of missile development and test programs, a vast amount of information is desired on such subjects as launching information, flight data, and automatic homing. It may be necessary to obtain all or part of the following information:

- Changes simultaneously involving roll, pitch, and yaw.
- Determinations of airspeed and altitude.

• Aerodynamic data obtained from various accelerations.

• Ambient conditions of temperature, humidity, and pressure.

• Structural information involving vibration and strain.

• Control data pertaining to the functioning of the control receiver, autopilot operation, servo operation, and/or displacement of control surfaces.

• Operation of homing or target-seeking equipment.

• Propulsion data including fuel flow and thrust.

• Ordnance functions, such as fuze-arming time.

• Upper air research (cosmic ray determinations, etc).

• Performance of electrical systems.

• Information regarding the operation of the telemetering equipment itself. Many of these measurements are interrelated. Some require a high order of time resolution, while for others a few samplings per second are adequate. Thus, a telemetering system must be capable of collecting, transmitting and recording large amounts of varied data per second.

Missile instrumentation may be broadly classified, with respect to methods, into two types of systems: *external* and *internal*. We'll take up external systems first, leaving the discussion of internal systems for sections II and III. Section IV covers direct-recording instrumentation.

Chapter 11 • Section 1

External Telemetering Methods Used in Guided Missiles

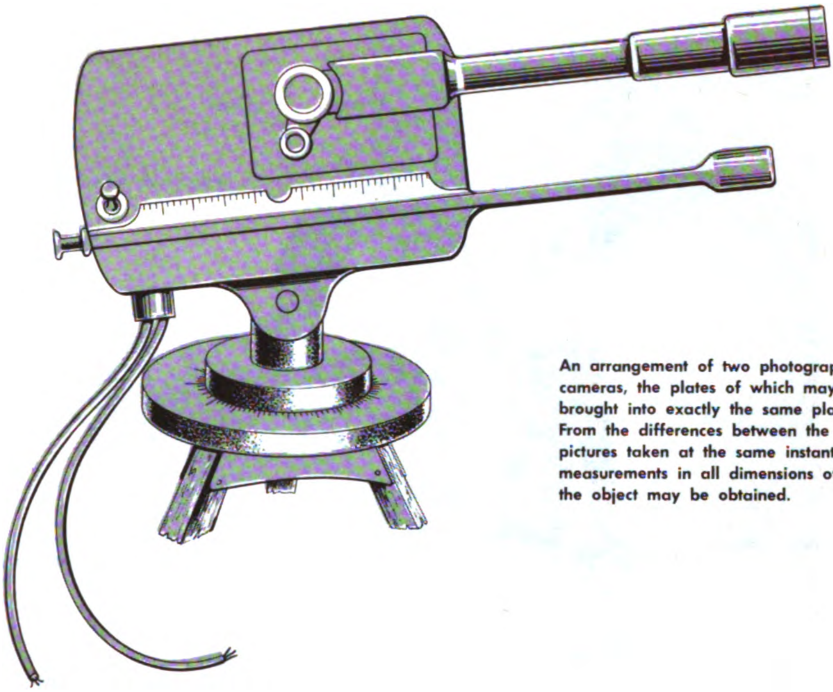
External systems involve the use of equipment for observing missile operations by optical or radar systems which employ no components within the missile itself. An exception to this is the use of a beacon transmitter in certain missiles to aid the ground radar equipment in locating and tracking the missile.

OPTICAL EQUIPMENT USED IN EXTERNAL INSTRUMENTATION

Optical instrumentation employs phototheodolites and synchronized telescopic cameras, or similar devices, to visually observe the missile along its flight path. With such equipment, information is obtained as to the missile's azimuth, elevation, height, range, and attitude, and as to the position of its control surfaces at any instant within visual range.

The ground station of an external system is equipped with visual observation and recording equipment such as "Askania" phototheodolites, by which flight characteristics of a missile or aircraft are observed and photographed over considerable distances. When phototheodolites are operated in conjunction with the FM/FM or radar telemetering equipment and are accurately timed, photographic data is obtained. The photographic data is coordinated with the radio data in order to obtain additional details.

Basically, a phototheodolite is an arrangement of two photographic cameras, the plates of which are brought into exactly the same plane. From the differences between two pictures taken at the same instant, measurements in all dimensions of the object can be obtained.



An arrangement of two photographic cameras, the plates of which may be brought into exactly the same plane. From the differences between the two pictures taken at the same instant, measurements in all dimensions of the object may be obtained.

A phototheodolite is a composite of two cameras

Phototheodolites operate on the principles of standard cameras with long-range telescopic lens systems that function in conjunction with standard components. Note the illustration above.

Each camera is a motion-picture-type camera with its shutter mechanism accurately synchronized with the electronic telemetering system by a common timing-control unit.

A theodolite consists essentially of a telescope mounted on standards so that it can be rotated vertically. The horizontal axis of the vertical rotation passes through the center of a vertical circle scaled for elevation-angle measurements.

The standards by which the telescope is supported are mounted on a calibrated base plate which can be rotated horizontally through 360°. The vertical axis of the horizontal rotation passes through the center of a horizontal circle which constitutes the azimuth scale of the theodolite.

Open sights, bubble levels, micrometric vernier scales, tangent screws, and scale-illuminating lights all add to the accuracy and versatility of the instrument.

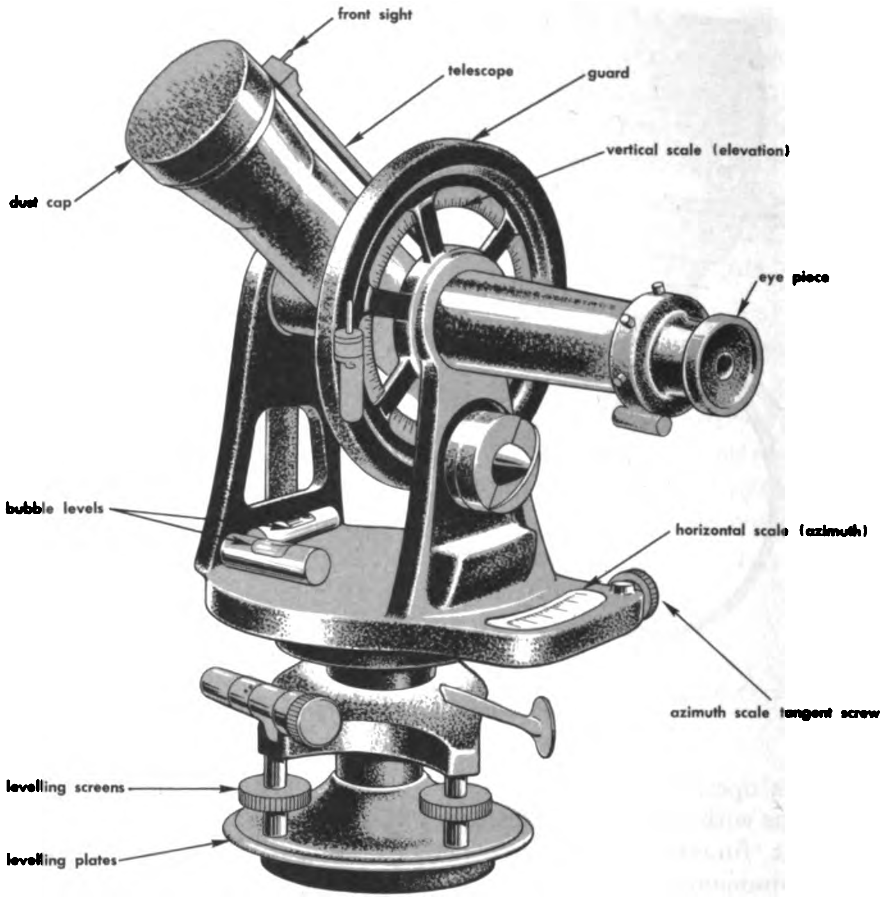
A typical theodolite as used by civil engineers and geodetic surveyors is shown in the illustration on next page.

Theodolite telescopes generally are of approximately 20 power and have a narrow field of view (about 2°). Cross hairs are provided for accuracy, centering the optical system on the object. These instruments are constructed with extreme care and possess a high degree of accuracy; consequently, they must be handled carefully and protected from heat, moisture, and severe shock to prevent impairing their accuracy.

At a missile test range, two or more theodolite camera stations are set up at suitable points along the range course. These stations are supplied with timing signals from a central control system so that the photographs taken during the missile flight can be identified and coordinated (telemetered) with other flight data for any given instant. The illustration on page 491 shows such a setup.

Flight data obtained from a theodolite camera (phototheodolite) setup includes:

1. Range data (distance).
2. Acceleration data.



Typical theodolite used by civil engineers and geodetic surveyors

3. Velocity data.
4. Trajectory and altitude data.
5. Explosion data.
6. Attitude data.
7. Booster separation and ignition data.

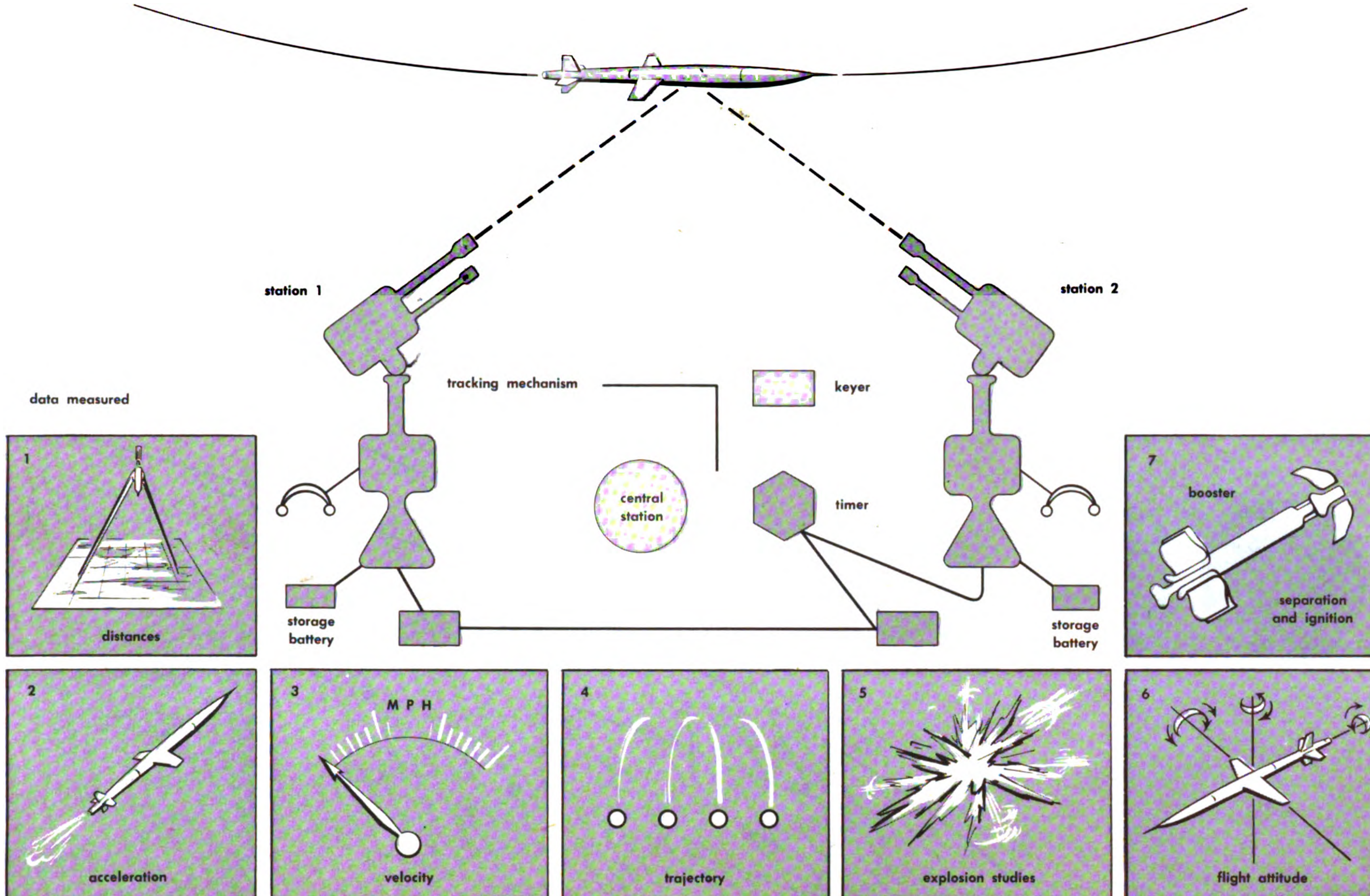
Additional details are supplied by these photo records, which supplement the graphic data obtained from the recording oscillograph and which serve as a check on the accuracy of the telemetering system. A telemetering system is pictured on page 492.

During long-range flight observations, more than one optical observation station is required. For accurate data determination, the ground stations must be synchronized so that the data collected at separate stations can be consolidated into a single *continuous* recording.

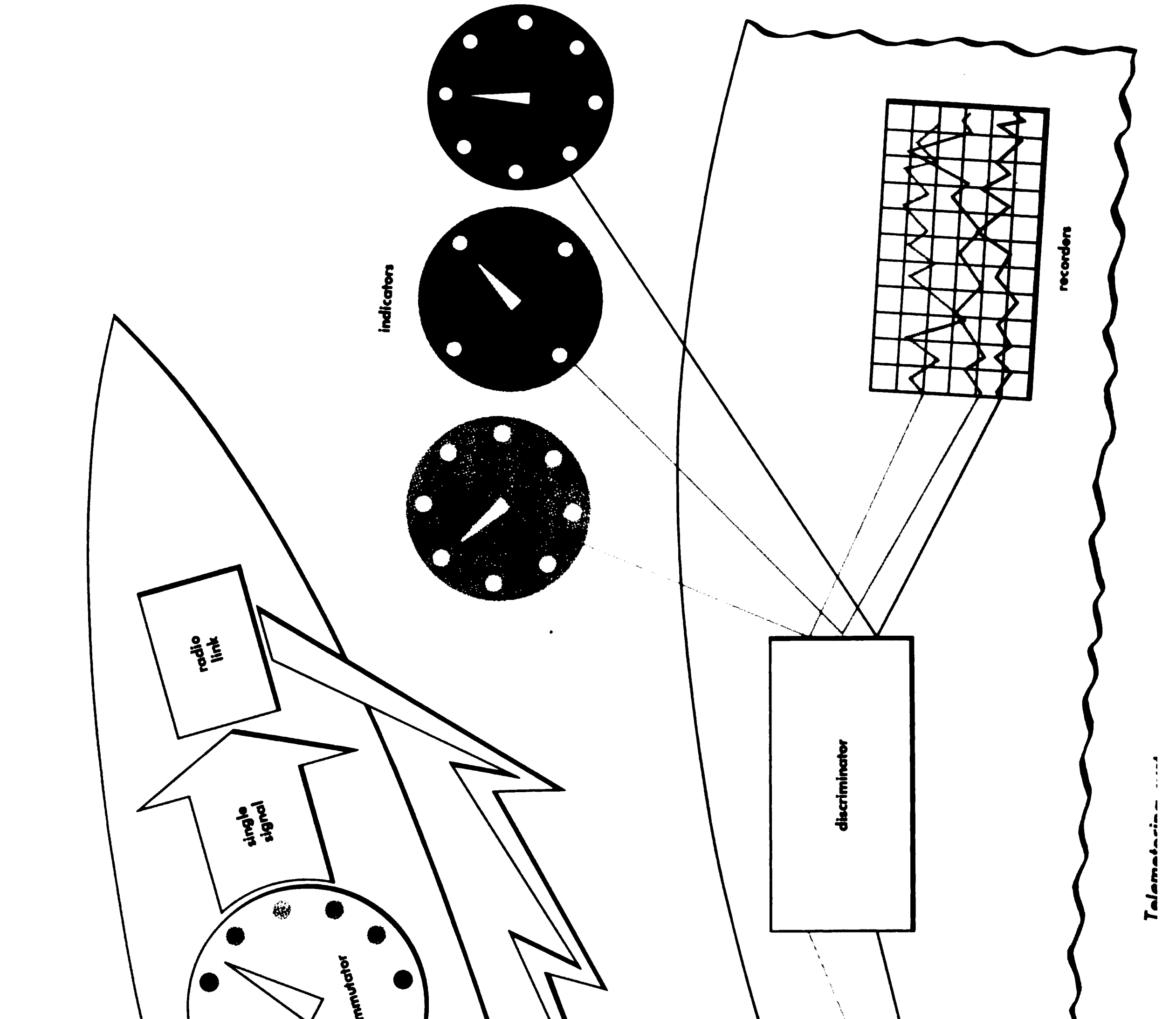
RADAR EQUIPMENT USED IN EXTERNAL INSTRUMENTATION

External radar instrumentation of missiles requires a radar *tracking and plotting system* capable of tracking the missile through its flight path and capable of presenting azimuth, elevation, height, and range position at any instant. Within the tracking and plotting system, there must be a data-transmission system to convey the tracking information to a computer and/or plotter. The tracking data is resolved into a permanent recording of the flight path and flight characteristics of the missile by the computing and plotting system. To do its job, a *radar tracking system* requires:

1. A radar transmitter to produce high-powered pulses of RF energy.



Theodolite camera set-up


indicators
recorders
radio link
single signal
commutator
converter
end instrument
pickup
pickup
pressure
temp
stress
missile movement
discriminator
selector
receiver

2. An RF system to conduct the pulses to the antenna and radiate them into space as a narrow beam. The RF system also picks up and concentrates the echoes (reflected pulses) from the target missile and conducts them to the receiver.

3. A receiver to detect and amplify the echo signals.

4. A range system to time the operation of the radar by measuring the time interval between the transmitted pulse and the echo signal, thus determining range information which can be presented in visual form on a *range scope*.

5. A PPI (plan position indicator) scope to present target range and azimuth data visually.

6. An antenna-positioning system to control the position of the antenna and indicate target azimuth and elevation on dials or suitable indicators.

7. A data-transmission system to supply instantaneous target data to a computer or plotting board so that the flight data of the missile being tracked (observed) can be analyzed and recorded.

Radar systems utilized in missile instrumentation generally employ automatic tracking antennas which use some form of *conical scanning* to develop the positioning data voltages required for their operation. When the target is not in the center of the beam, *error voltages* proportional to the deviation in elevation and azimuth are developed in the antenna-positioning system.

These error voltages, or *position-data voltages*, when compared in phase or polarity with a reference voltage representing a fixed azimuth or elevation angle of the antenna, provide the azimuth and elevation data which indicate the instantaneous position of the target missile with respect to the radar antenna.

Typical Radar Set Used for Instrumentation

A typical radar set which has been used in a modified form in missile instrumentation is the SCR-584. This set, as shown in the diagram on the next page, is divided into the following seven systems: transmitting, radio-frequency, receiving, range, PPI, antenna, and data transmission.

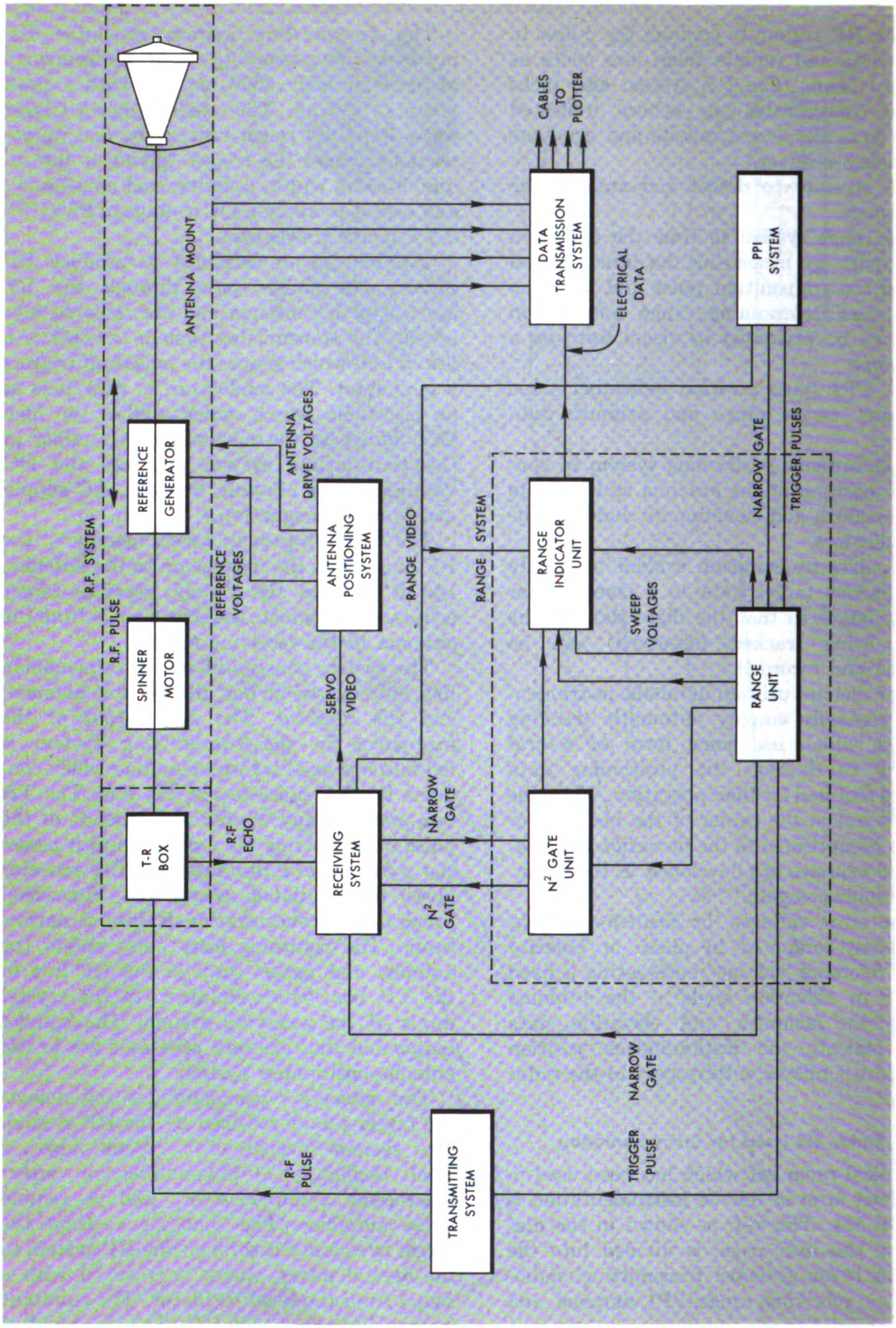
The transmitting system generates RF pulses 0.8 microsecond (μ) wide at a frequency of approximately 2800 mc and with a peak power of 210 kw. The system takes a trigger pulse from the range unit every 586 microseconds, shapes the trigger (giving it the required form, width, polarity, and amplitude), and uses it to apply a DC voltage of 22 kv for 0.8 μ to the magnetron.

The magnetron oscillates to produce RF pulses. The pulses pass through the RF system to the antenna and are radiated into space. The transmitting system consists of a *driven unit* which shapes the pulse and triggers a modulator. The modulator, in turn, acts as an electronic switch which applies the high DC voltage to the magnetron. A rectifier in the transmitting system produces the DC voltage from a 115-volt, 60-cycle, AC supply and from the magnetron.

The *radio-frequency system* conducts the RF pulse from the magnetron to the antenna, and it radiates RF energy, receives target echoes, and conducts the echo signals from the antenna to the receiving system.

The system consists of an RF transmission line, antenna switch box, the TR box, antenna, and the reflector. The RF output of the magnetron in the transmitting system is fed into a coaxial transmission line which conducts the RF pulse to the TR box. The TR box prevents any appreciable amount of the pulse from entering the receiver, and it passes the pulse toward the antenna. The reflector gives this radiated energy its directional property by concentrating it into a narrow beam. The returning echo pulse enters the antenna and passes through the RF line to the TR box, which shunts it to the crystal mixer of the receiving system. The spinner motor and the reference generator are in the antenna-positioning system.

The *receiving system* amplifies and detects the target echoes received by the RF system. The system consists of a crystal mixer, a local oscillator, a preamplifier, a superheterodyne radio receiver, and a remote video amplifier. The echo signal from the target is conducted through the RF system to the crystal mixer, where it is mixed with a signal from the local oscillator. The resultant intermediate-frequency (IF) signal is amplified



Typical radar system (SCR-584) used in tracking and plotting missile instrumentation

in the preamplifier and the IF section of the receiver.

The IF signal then divides and enters the range and servo channels of the receiver. The signal is amplified in the range channel, then fed to the scopes in the range and PPI systems. The servo channel is gated by the pulse from the N^2 (narrow-narrow) gate unit. As a result, the channel is opened up only long enough to receive the echo from the target being followed. In the servo channel, the error voltages corresponding to the echo being tracked are passed on to the antenna-positioning system.

The *range system* furnishes the timing and gating pulses and visual range data on oscilloscopes. The system also supplies range potentiometer and selsyn data to the data-transmission system. The range system consists of a range unit, a range indicator unit, an N^2 gate unit, and a range power supply and rectifier unit.

The range unit generates the trigger pulses and sweep voltages for the set. It furnishes synchronized triggers to the transmitting system, the PPI system, and the other components of the range system. It also supplies a narrow gate to the PPI system as well as sweep voltages and illuminating gates to the range indicator. Too, the 82-kc master timing frequency and the narrow gate are supplied to the *N^2 gate unit* by the range unit.

The *range indicator unit* consists of two range scopes, *tracking* and *slewing* handwheels, and an added tracking mechanism. The range scopes are *J-scan* (circular sweep) scopes; their sweep voltages are generated in the range unit. All targets within range of the RF beam are displayed on the 32,000-yard scope (coarse range) which is illuminated for the duration of the wide gate. The 2000-yard scope (fine range) expands any 2000-yard portion of the 32,000-yard scope for greater tracking accuracy. Two gates are applied to the 2000-yard scope: the narrow gate and the N^2 gate. The narrow gate illuminates a sector on the 2000-yard scope, which may be manually varied by the narrow-gate width control from 250 yards to 1800 yards. The N^2 gate is applied to the 2000-yard scope so that the operator may track a specific target. Range potentiometers and selsyns supply *slant range* data to the data-transmission system.

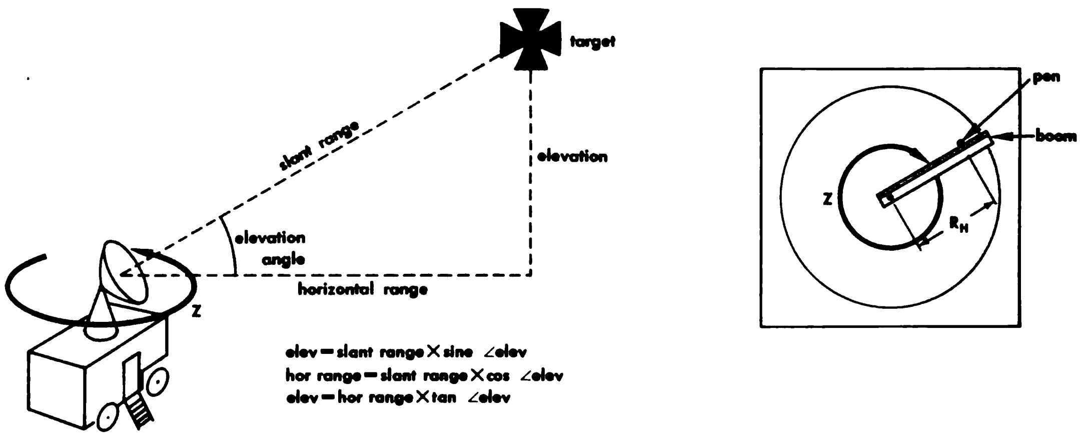
The N^2 gate unit generates the 50-yard wide N^2 gate from the 82-kc master timing frequency after being triggered by the narrow gate. This N^2 gate is fed to the servo channel of the receiver and also to the range indicator. It appears on the 2000-yard scope as a slight increase in intensity, 50 yards wide, and can be seen at the top of the echo signal being tracked. The range operator tracks the target by keeping the range hairline at the leading edge and the N^2 gate at the center of the target being tracked.

The PPI system is used for general searching and provides the means of initially locating the target. The system includes a PPI unit, the PPI scope, the PPI power supply and rectifier, and the PPI selsyn.

All targets within the RF beam are displayed on the PPI scope together with the narrow gate and range-marker circles. Thus, the PPI scope furnishes information for positioning the tracking hairline of the range scope. By the proper adjustment of handwheels, the target is placed within the tracking gate. The sweep is triggered by a pulse from the range unit. The sweep line rotates with the antenna in azimuth, enabling the PPI scope to indicate the azimuth direction in which the antenna points.

The *antenna-positioning system* controls the positioning of the antenna. This system includes the antenna-position-control unit, the automatic tracking unit, the azimuth and elevation tracking unit, the azimuth and elevation motor-generators, the spinner motor, and the reference generator (two phase).

In *automatic tracking*, the error voltage is combined with the reference voltages, thus producing the desired control voltages for the drive motors. When positioning the antenna manually, *artificial* error signals, created by moving the azimuth and elevation handwheels, are combined with a reference voltage to produce the same result. A single group of circuits serves for all methods of antenna control, with the difference in the operation of the circuits depending on the nature and source of the error and reference voltages.



Basic RC-294 plotter

The *data-transmission system* converts the range, azimuth, and elevation information on target position into potentiometer (voltage) and selsyn (phase-position) data, and the system supplies the data to plotting boards. The system includes an antenna-position indicator; azimuth, range, and elevation potentiometers; selsyns; an attitude data unit; and an altitude converter. *Coarse* and *fine* selsyns transmit angular data for azimuth and elevation. Potentiometers develop range, altitude, and angular data for azimuth and elevation. The selsyns and potentiometers respond instantly to all changes in target position. Transmission of data is electrical through the use of control voltages developed across the potentiometers, or mechanical by means of the selsyns. Accuracy of the data supplied to the gun director or plotter depends primarily on the accuracy with which the set tracks the targets.

The mechanical (selsyn) and electrical (potentiometer) data obtained from the tracking radar is resolved into "X" (range) and "Y" (elevation) components. These components are proportional to the variations of the angular error voltage, "Z" (azimuth angle) data, with respect to the reference voltages. The "X" and "Y" components are converted into graphic form in the plotting system.

Basic Plotting System

An example of a basic plotting system is the RC-294 plotter, illustrated above, in which the plotting boom revolves around a center axis and is positioned by the azimuth (Z) data. The tracing-pen trolley is positioned by horizontal range data. The axis of the boom represents the radar location, and the pen position indicates the instantaneous location of the target. The horizontal range data used to position the tracing-pen trolley is derived from the slant range data in terms of the cosine of the elevation angle. Elevation data is in terms of the sine of the elevation angle and slant range data or in terms of the tangent of the elevation angle and horizontal range. If desired, the data is marked on the plotting chart at each target-position point.

Many improvements and modifications have been made in the basic radar tracking and plotting equipment to meet the requirements of the guided missile program. Automatic tracking over full range of the radar and greater flexibility of the antenna-positioning system have been provided in some modifications. Larger and more versatile plotters with integral auxiliary indicating and recording devices have been developed to meet the requirements for tracking high-velocity, long-range missiles. However, the same principles of operation prevail in the new modified equipment as are embodied in the basic radar tracking and plotting systems.

You can see that, as a rule, external methods of missile instrumentation do not involve using components within the missile. Now let's consider in the next two sections the methods of instrumentation that do involve using components within the missile.

Internal Telemetering Methods Used in Guided Missiles

An internal system utilizes pickoff devices and a data transmitter located within the missile, and receiving and recording equipment located in a ground station. Internal systems are capable of supplying a much greater variety of data than can be obtained from the external systems, and are not so restricted by conditions of weather, visibility, or terrain.

Whenever it is desired to collect data of a physical or mechanical nature and transmit it from a missile to a ground station, the data must be converted into an electrical form. By making this conversion, the data can be transmitted in the form of modulation of an RF carrier or transmitted as radar pulses by a transmitter located within the missile to a receiver located at the ground station.

This process is referred to as *telemetering*. The word is of Greek origin. It means measurement from a distance, and it may be defined as a method of measuring a quantity, transmitting the result to a distant station, and there indicating or recording the quantity measured. Telemetering includes conversion of quantities to be studied into electrical signals, the transmission of these signals over a radio link, their reception, and their presentation by indicators or permanent recordings. The quantities in this final form of presentation can be studied at leisure.

Another possible means of gathering data is through photographic or other recordings that are made aloft, but this method is not considered telemetering since there is no distance involved between the instruments and the recorders. This method has proven impractical for use in a missile because of the doubtful recovery or photographic- or tape-recorded data after the missile has expended itself. Photographic and tape recordings are discussed in detail in Section IV of this chapter.

The use of television for gathering data is also ruled out when space and weight are considered.

In missile telemetering, some type of radio transmission has proven most practical. Present-day systems generally employ one of the following types of transmission:

1. FM/FM (frequency-modulated/frequency modulation).
2. AM (amplitude modulation).
3. PM (phase modulation).
4. Radar (pulsed radar-time sharing).
5. Television.

END INSTRUMENTS

Internal instrumentation of missiles includes the use of data pickups or *end instruments* which are attached to, or form an integral part of, the component under test within the missile.

These end instruments, if not part of an electrical system, are supplied with a reference-voltage source which is preset to some output value corresponding to the normal position or operating characteristic of the component

under test. Any deviation from *normal* provides an output voltage (data) which varies from the *normal* value by an amount corresponding to the amount of the component variation.

An end instrument measures or *collects* the data at its source. The nature of the data to be picked off from the source determines the type of pickup or end instrument employed. The terms *pickup*, *pickoff*, *end instrument*, and *transducer* are used more or less interchangeably throughout the missile field.

The end instrument that accomplishes conversion of mechanical or physical variations into corresponding electrical variations is called a *transducer*. A transducer converts these variations into a desirable voltage. This voltage is used to modulate the carrier to a degree corresponding to the degree of change in the quantity being measured. If an electrical circuit is being *sampled*, it may be possible to obtain data voltages within the required limits by direct circuit connections.

Whenever possible, an end instrument is built into the component under test as an integral part to avoid external loading, drag, or shunt effect of the pickoff device on the component.

Small, light-weight, highly accurate end instruments have been especially designed for telemetering. A few of these instruments which are commonly employed in missile telemetering systems include:

- Potentiometers (rudder-position data).
- Strain gauges.
- Thermistors.
- Selsyns (roll signal).
- Tachometers with integral E coils (engine rpm).
- Thermocouples (nozzle temperatures).
- Thermostats.
- Manometers.
- Bourdon tubes (airspeed).
- Diaphragms or bellows.
- Direct connection (dive signal).
- Vanes.
- Gyros.

As the speed of a missile increases and reaction times become less, accurate time studies become more important. When you consider that the entire time devoted to collecting data is only a small portion of the time taken for the entire flight, the seriousness of the problem of obtaining accurate data becomes readily apparent.

The problem starts with the end instruments, of which there are many varieties. The more common types depend on either motion, temperature, or current changes. Such end instruments as diaphragms, bellows, bourdon tubes, gyros, autosyns, and vanes depend on motion for their action. These take in a large variety of measurements including altitude, pressure, position, fuel flow, etc. Force, pressure, velocity, and acceleration measurements usually involve mechanical motion.

Even with strain-gauge measurements, there must be motion, however slight. Whenever there is motion, the mechanics of the system, including damping and inertia, become a factor. The sensitivity generally varies inversely as the square of the frequency; that is, a sensitive system, is sluggish, and a system with a more rapid response is less sensitive to small changes.

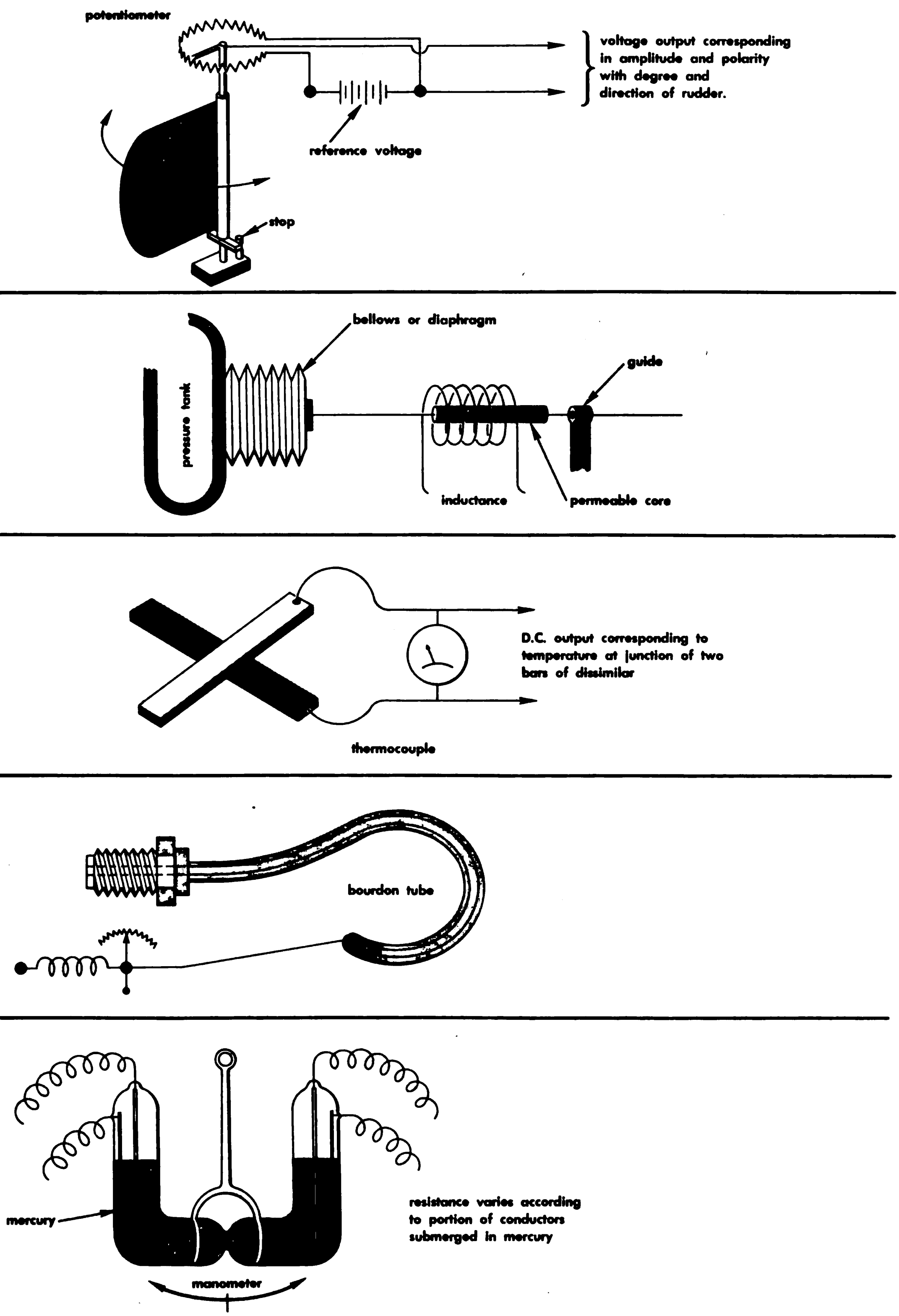
The time resolution and accuracy of telemetering channels need not exceed the time resolution and accuracy of the end instruments. The time resolution and accuracy of telemetering may also suffer in the recording process. This could be due to the optics of the recording means or perhaps variations in speed at which film or paper travels through the recorder.

Let's consider a few types of end instruments.

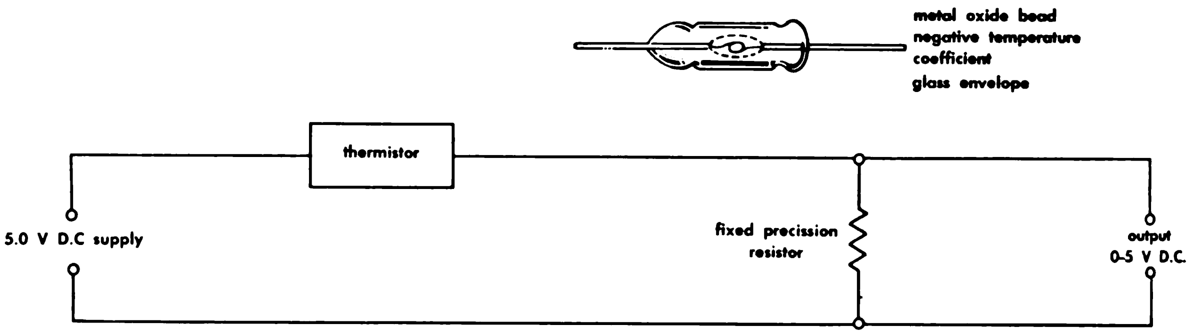
Types of Transducers

A transducer is generally considered to be a data-transfer unit which includes an end instrument and a pickoff device. It is just as important that the transducers be able to withstand rigorous conditions of acceleration, vibration, and temperature as it is for any other link in the system.

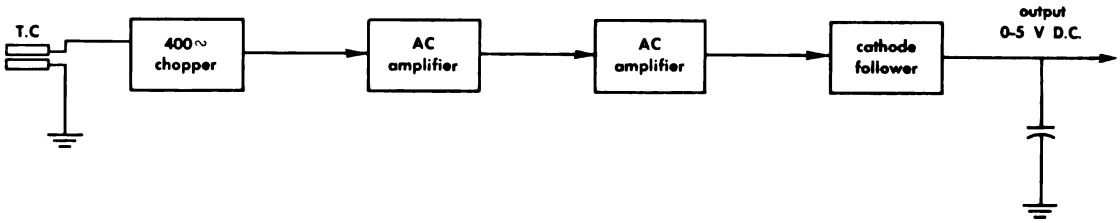
The many transducers in use at present preclude more than a general coverage of the field. They may be classified according to whether they are *modulating* or *generating* in action. Examples of the modulating type of



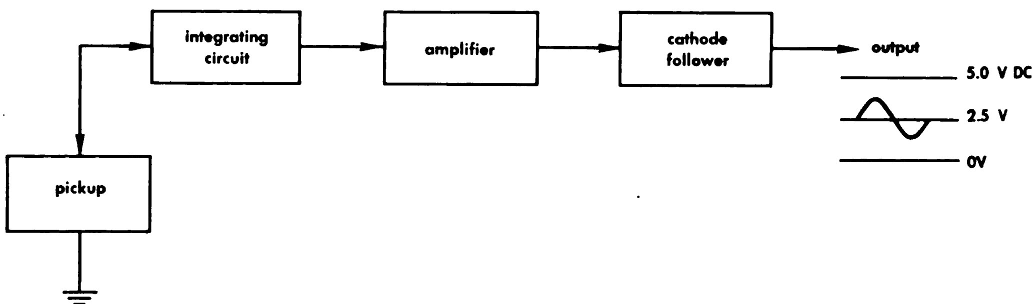
Common end instruments



Thermistor pickup circuit



Thermocouple amplifier transducer unit

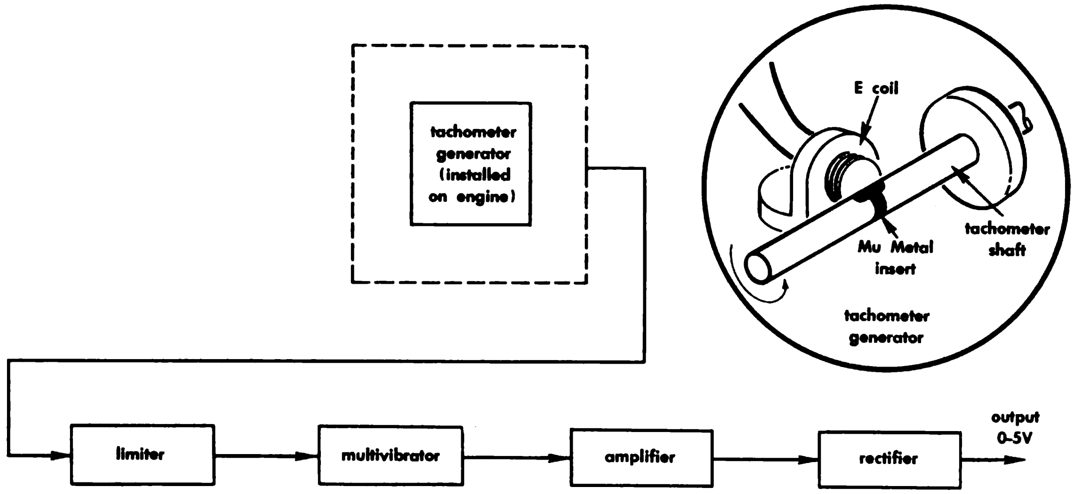


Vibration amplifier transducer unit

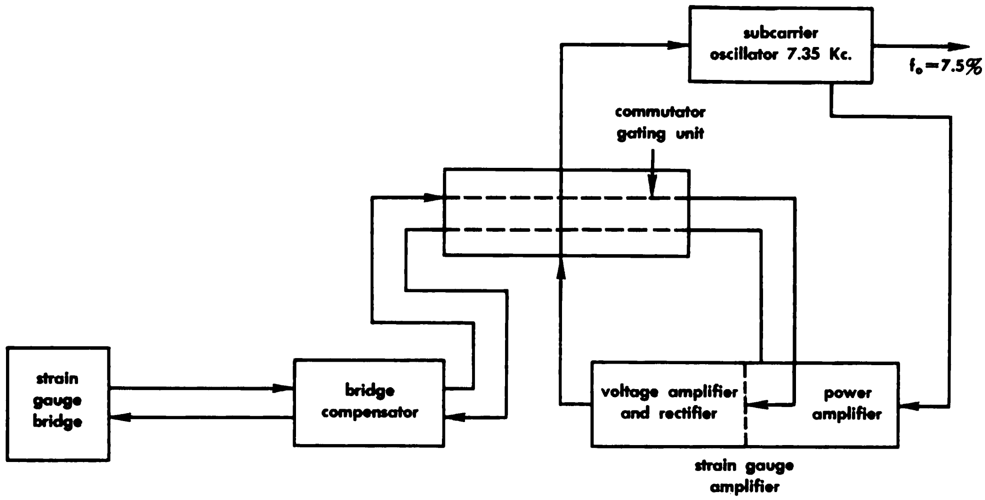
transducers are the variable inductance, variable capacitance, and variable resistance. Variable-resistance transducers include potentiometers, resistance strain gauges, thermistors, and electron tubes.

Generating types of transducers are photoelectric, thermoelectric, piezoelectric, and sometimes electron tubes. The generating type, as applied, is not always capable of completely modulating the designated circuit of the transmitter, nor is the modulating type always capable of completely performing this function. In such cases, their outputs are further amplified. Examples of required stages associated with these pickup devices are shown in the accompanying diagrams of transducers.

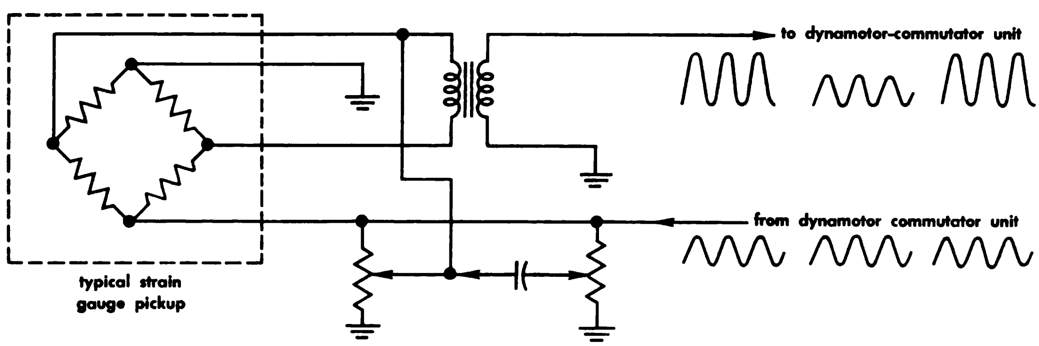
A primary step in instrumenting a missile is to calibrate the instrument with its transducer attached. Often, standard aircraft instruments are modified by the addition of the transducer. In such cases, the transducer must not impair the operation of the instruments. If close indication of a certain intelligence is desired, an instrument and transducer which are combined into a unit designed for the specific purpose provides the greatest accuracy.



Tachometer amplifier unit



Commutated strain gauge system



Bridge compensator

Types of transducer units

Another method is to couple a variable-reluctance type of transducer to the instrument by attaching a small piece of magnetic material to the needle or moving part. The torque then necessary to drive the needle is extremely low. This setup is best fitted to work into a frequency-changing arrangement such as results when the reluctance of a frequency-controlling inductance is caused to vary with motion. Usually, this is the tank coil of an audio-frequency oscillator. It is also possible to obtain the change in frequency from a large change in air gap with increased spacing of the metal. Such transducers are excellently adapted for attachment to such instruments as accelerometers, altimeters, airspeed indicators, pressure gauges, tachometers, and various displacement meters.

One type of transducer easily adapted for use with tachometers consists of a segment of *mu metal* (metal of very high permeability) imbedded in the rotating shaft of the instrument and so located that as the shaft rotates the mu metal moves in and out of the field of an inductor which is a component of an oscillatory circuit. An example of this type of transducer is shown below.

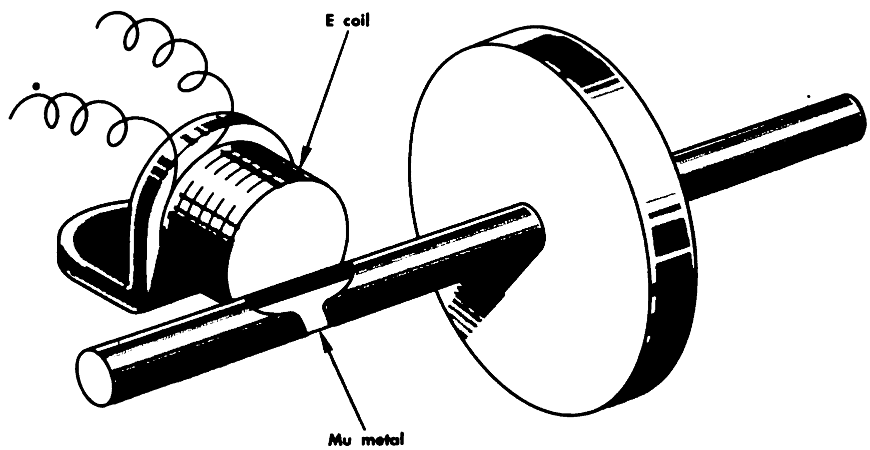
This arrangement varies the inductance of the coil and produces a change in the frequency of the oscillatory circuit at a rate corresponding to the rpm of the shaft. Mu metal is used in this application because of its freedom from hysteresis (lag in magnetic change), which is common in most magnetic materials and which might introduce errors.

The mu metal and E-coil transducer offers little drag on the rotating device. It can be used as a miniature alternator whose frequency and amplitude of output voltage are dependent on the rpm of the shaft and the spacing between the metal segment and the coil.

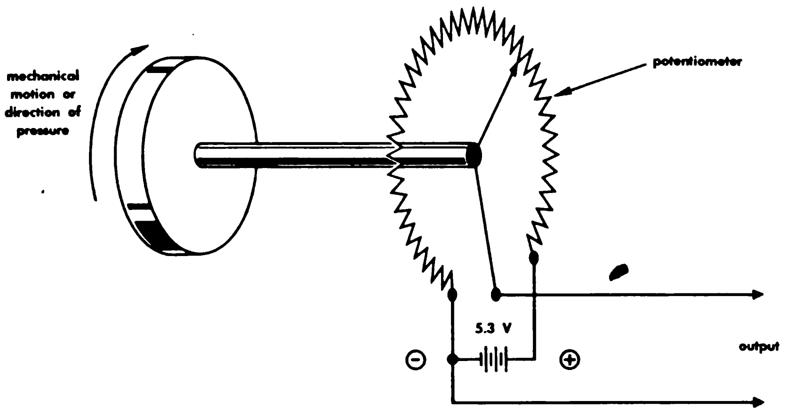
Another typical transducer assembly consists of linking the mechanical-data source to the wiper arm of a linear potentiometer on the order of 2000 ohms, accurate from +10% to 0%, which is supplied with a reference voltage from a 5.3-volt reference battery.

In operation, an output of 5.3 volts represents maximum mechanical or pressure action, and an output of 0.0 volts represents minimum action. As mechanical action is initiated or pressure developed, the output voltage rises from 0.0 volts to a positive voltage in proportion to the mechanical action, maximum action resulting in +5.3-volt output.

The entire assembly of potentiometer and mechanical analyzer, illustrated in the figure above right, is inclosed in a sealed container.



Transducer adaptable for use with tachometers



Transducer unit

Transducers are used as end instruments in missiles for the following telemetered data:

- Altitude
- Static differential pressure
- Low airspeed
- High airspeed
- Vertical accelerometer
- Electronic chamber pressure
- Hydraulic pressure
- Plenum-chamber pressure
- Nozzle pressure
- Slip and pitch
- Surface position (stabilizer-spoilers)

Thermistors

Thermistors, consisting of combinations of ceramic materials and various metallic oxides, are commonly used to detect and respond to fluid- or air-temperature changes. A hydraulic fluid-temperature thermistor is incorporated in an AN-type plumbing plug which withstands gauge pressures up to 2250 pounds per square inch and has a temperature range from −40°F to +250°F, resistance versus temperature being linear (70K ohms at −40°F to 2.1K ohms at +250°F).

Air-temperature thermistor elements, generally located on the center and aft equipment shelves in a missile, are mounted in such a manner that air can circulate freely around them but so that they are protected from external damage. The temperature range of an air-type thermistor is from −40°F to +160°F, with resistance versus temperature being linear (100K ohms at −40°F to 5.5K ohms at +160°F).

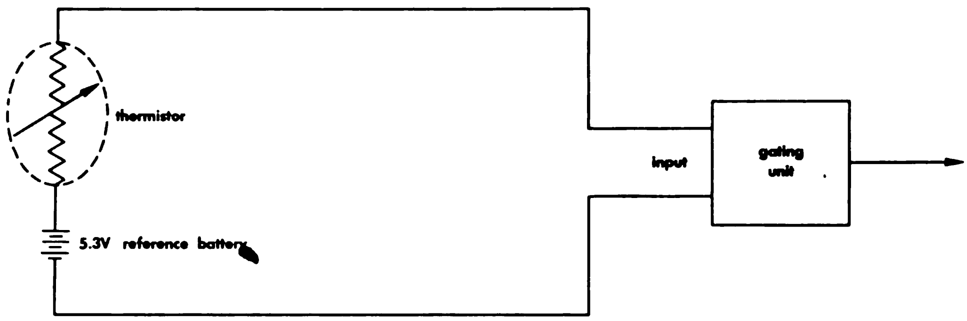
The thermistor element is connected in series with a 5.3-volt reference battery and its assigned gating-unit contact. The top diagram on the following page shows a thermistor circuit and its connection to a gating unit. As the thermistor temperature varies, the resistance varies; therefore, the voltage output to the gating unit is linear with the temperature variations. An increase in temperature decreases the resistance of the thermistor, producing a smaller voltage drop across it. This action, in turn, causes a higher voltage to appear across the unchanging impedance of the input to the gating unit.

Strain Gauge

Strain gauges are variable-resistance elements so designed that any mechanical movement or stress applied to them causes a corresponding variation in their resistance. They are made in many forms and types to suit application requirements. A common type of strain gauge consists of an adhesive *tab*, shown on the following page, in which one or more loops of fine resistance wire are imbedded. Connections are the same as for thermistors.

FM/FM TELEMETERING SYSTEMS

The most widely used telemetering system currently employed in missile instrumentation is the frequency-modulated subcarrier, frequency-modulated carrier type, referred to as FM/FM.



Thermister circuit connected to a gating unit

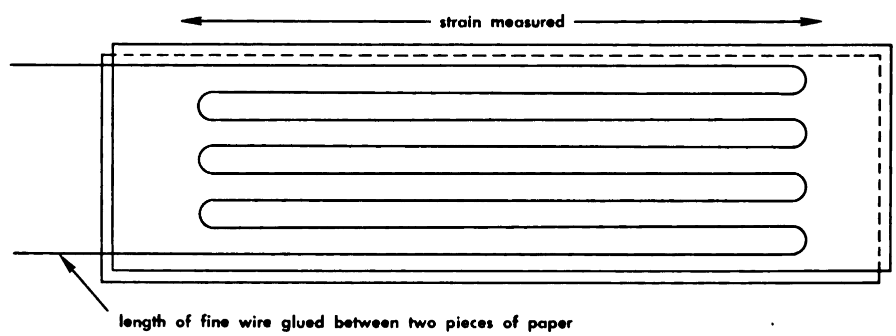
This system is commonly used because it can be designed with inherent freedom from external signals (noise) which are amplitude variations. Also, conventional equipment and circuitry can be used; space, weight, and power requirements are not extreme. Ruggedness and simplicity are attainable with a relatively high degree of accuracy and adaptability to missile telemetering requirements. Ease of calibration and a minimum amount of ground-station equipment are important factors, as well as versatility in handling many different types of data.

The underlying theory of operation is similar to FM broadcasting; consequently, specialized training of personnel does not present serious problems. Frequency modulation of the subcarriers also has an advantage in minimizing interference from other electrical circuits within the missile. In an FM/FM system, amplitude variations from the data pick-offs are converted into frequency variations by the transmitting system. These frequency variations are converted into voltage data of varying amplitude in the receiving system. The frequency deviations of an FM carrier are not affected by noise or static from sources outside the system because such types of interference, while tending to vary the amplitude of the carrier, affect its frequency very slightly.

In designing an FM/FM telemetering system, emphasis is placed on frequency stability of the subcarrier oscillators even though ambient conditions such as temperature, humidity, acceleration, or power-supply voltage vary as the flight progresses. It is important to remember than any drift in subcarrier frequency would appear as a change in the function being telemetered.

A typical FM/FM telemetering system includes the following airborne components:

1. A radio transmitter unit.
2. Subcarrier modulator units (one or more).
3. A commutator.
4. An antenna and antenna coupler.
5. A time-delay relay.
6. A dynamotor.



Strain gauge

7. A filament voltage battery (lead-acid type).
8. Junction boxes, cables, plugs, etc.
9. Gating unit.
10. Frequency-converter units.
11. End instruments.

The ground station for an FM/FM system includes:

1. Receiver unit (with subcarrier separation filters, subcarrier discriminators, recording oscillograph, and panoramascope).
2. Power supply.
3. Antenna.
4. Cables and connectors.
5. Recording meters.
6. Counters and computers.
7. Test equipment and calibrators.
8. Maintenance equipment.

When an airborne transmitter is required to collect data from a large number of sources in the missile, additional subcarrier modulators can be added to the transmitter to accommodate additional data channels. For data channels requiring *continuous sampling* or a high degree of time resolution, several additional subcarrier modulators are needed. For data of the slowly changing type, a single subcarrier modulator with its input commutated from a number of end instruments or channels may suffice. Additional commutators and subcarrier modulators can be added if required. If any are added, additional subcarrier filters and discriminators must be provided at the receiver.

If a sufficient number of research missile are available, the best practice is to telemeter a small number of channels, thus insuring the highest degree of accuracy. When many data sources must be sampled from a single missile the possibility of interference by cross-modulation in data channels necessitates careful shielding of cables and components and careful calibration of both transmitting and receiving equipment. Even so, the varying loads on the power-supply and transmitter components may introduce false indications.

Generally, when feasible, certain types of data are collected from one missile and other types from additional identical missile. Certain data may be in simple form requiring few samplings to produce a complete record of operations under various flight conditions.

Before a missile reaches the stage of development at which it can be put into production, and consequently into combat, it becomes necessary to launch many experimental models. It is from these *one-shot* test missiles that scientific information is gained through the avenues of telemetering.

Many types of intelligence are gathered when a test missile is fired; and after careful evaluation of this data, specific types of intelligence are filtered out, analyzed, and either set up for repetition or replaced by more desirable information in the next missile.

It is obvious that the number of different types of intelligence obtainable at one time is limited to the number of data channels available in the telemetering system being used. One FM/FM system widely used is provided with thirty-three data channels. However, from one to three of these channels are used for synchronizing signals to synchronize the commutator of the receiver with that of the transmitter so that data may be identified from the sequence in which it is sampled.

Some of the types of information that can be telemetered in missile research are:

1. Throttle setting.
2. Engine rpm (continuously telemetered).
3. Ambient temperatures (commutated data).
4. Compressor inlet temperature.
5. Nozzle temperature.
6. Structure temperature.
7. Altitude.
8. Airspeed (low).
9. Airspeed (high).
10. Induction air pressure (total).
11. Induction air pressure (static).
12. Nozzle pressure (total).
13. Nozzle pressure (static).
14. Fuel static pressure.
15. Fuel pressure (tank).
16. Fuel-flow rate.
17. Power-supply voltages.
18. Hydraulic pressure.
19. Groundspeed.

20. Generator voltage supply.
21. Airspeed signal input.
22. Yaw error (commutated).
23. Pitch error (continuously telemetered)
24. Roll error (continuously telemetered).
25. Missile control-surface position.
26. AGC. voltages.
27. Fuse arming.
28. Radio-control relays.
29. Terminal-dive signals.
30. Airspeed and altitude relays.

Some of the data above is relatively unimportant, and it has been deleted from the telemetering program. But about thirty types of data are still found to be essential and have been retained in the research program. New types of data pertaining to newer missiles developments are added from time to time.

Information to be telemetered is supplied to the FM/FM transmitting unit in the form of a DC voltage, varying in amplitude at the rate of change taking place in the function being telemetered. This voltage is obtained by passing the signal, if it is from an AC source, through a converter. If it is DC, it may be supplied directly from the end instrument.

Some end instruments used in FM/FM systems and examples of the intelligence they provide are:

1. Potentiometers — rudder position.
2. Selsyns — roll signal.
3. Tachometers — engine rpm.
4. Thermocouples — nozzle temperatures.
5. Bourdon tubes — airspeed.
6. Direct connection — dive signal.

FM/FM telemetering systems utilize an FM RF carrier modulated by a number of FM subcarriers. These subcarriers may be continuously modulated by varying information or by commutated (sampled) information or a combination of these methods.

In general, information from any of the three available types of pickups (potentiometer, strain gauge, variable reactance) is used to frequency modulate a number of subcarrier oscillators. The oscillator outputs are combined and amplified and then used to modulate the FM RF carrier of the transmitter which generally operates in the 215-230 mc range.

A transmitter with a nominal output of three watts can feed an antenna directly, or it can be used to drive an auxiliary RF amplifier, which raises the output to a nominal 30 watts.

The equipment is used for any application desired so long as its use is consistent with the design of the equipment. For example, pickups of any of the three available types can be used to frequency modulate subcarrier frequencies continuously with varying information. This permits the information from one pickup per subcarrier to be telemetered. Also, a number of subcarrier frequencies can be modulated using continuous information from pickups of any of the three types, while one or more of the higher subcarrier frequencies are modulated by as many as 27 additional commutated data inputs (each of a slowly varying nature) from either potentiometer or strain-gauge pickups.

Let's now consider a typical missile FM/FM telemetering system. A functional block diagram of such a system is presented on the opposite page. The discussion here is concerned with a specific system, but is applicable to other FM/FM telemetering systems in the missile field.

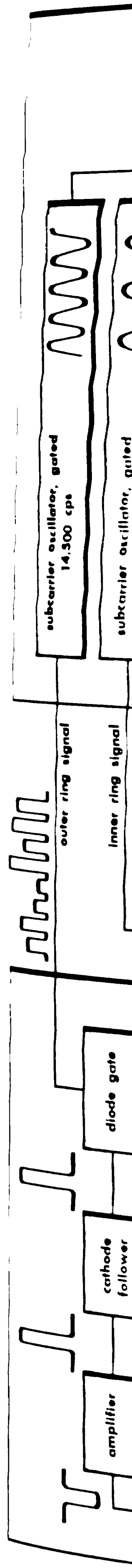
This telemetering equipment is designed to withstand accelerations of 10 G's in any axis and to operate satisfactorily at any temperature from $-10°C$ to $+75°C$ at relative humidities up to 95% and at air pressures equivalent to those that exist at altitudes up to 60,000 feet.

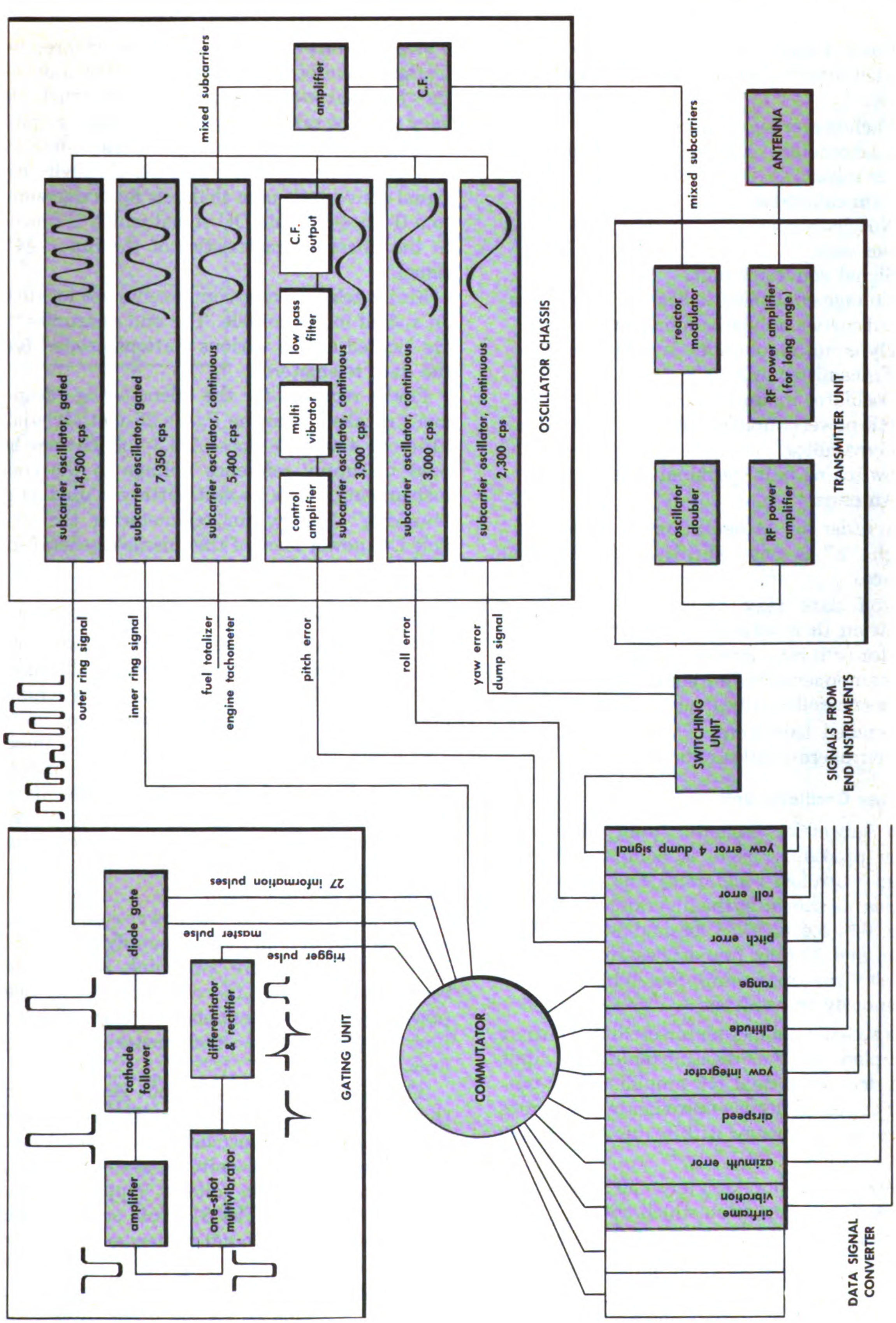
The FM/FM system described here is designed to transmit data on six subcarrier frequencies. The sixth (highest-frequency) channel is normally commutated at the rate of five samples per second for each of 27 data pickoffs. Remember, commutated pickoffs are used only where the information is expected to vary at a very low rate.

Channels are modulated continuously when the intelligence is expected to vary more rapidly or where a high degree of data resolution is required. Systems employing higher commutation rates are available if it is desired to commutate information varying more rapidly than one cycle per second.

The system that we're discussing includes the following components:

1. Subcarrier oscillator unit.





Functional diagram of an FM/FM telemetering system

2. Power-supply unit.
3. Fuel totalizer; engine-tachometer switching unit.
4. Tachometer signal generator.
5. Tachometer, low/high airspeed, fuel totalizer mixer.
6. Cam calibrator.
7. Nozzle-temperature thermocouples and amplifier unit.
8. Signal converter unit.
9. Voltage-controlled subcarrier oscillators.
10. Subcarrier oscillator mounting unit.
11. Dynamotor-commutator gating unit.
12. Transmitter unit.
13. Oscillator mounting unit.
14. RF power amplifier unit.
15. Dynamotor.
16. Switching unit (terminal dive).
17. Antenna.

The order of presentation of items "a" through "h" is not indicative of the signal sequence.

Pickoff data may be suitable for direct application to a voltage-controlled subcarrier oscillator or it may involve one or more of the other components before being applied to the voltage-controlled subcarrier oscillator.

A detailed treatment of these components and their operational functions follows.

Subcarrier Oscillator Unit

The subcarrier oscillator unit, illustrated on the preceding page, is composed of six voltage-controlled oscillators. The center frequencies, as shown in the diagram, are 2.3 kc. 3.0 kc, 3.9, 5.4 kc, 7.35 kc, and 14.5 kc. The 7.35-kc and 14.5-kc frequencies are the *gated channels;* the remaining four channels are continuously transmitted.

The six-channel modulator-oscillator output is coupled to the transmitter/exciter unit which drives the final RF amplifier. The output of the amplifier is approximately 30 watts at a center frequency within the 215-230 mc range.

There are approximately 50 points of telemetering information which could be used on each missile. The average information telemetered per missile is 30 input signals to the system.

In this FM/FM system, the data signals of 0 to +5.3 volts amplitude are coupled directly to the dynamotor-commutator-gating unit or the modulator-oscillator. Data expected to vary at a low rate is coupled to the *gating* unit. AC signals received from the missile controls and guidance units are coupled through the signal converter's nine channels for conversion to a 0- to +5.3-volt DC signal which is linear to the phase characteristics of the input AC signal.

Major telemetering components are located on a shelf in the missile. The end instruments are located at the various stations needed for the data telemetered.

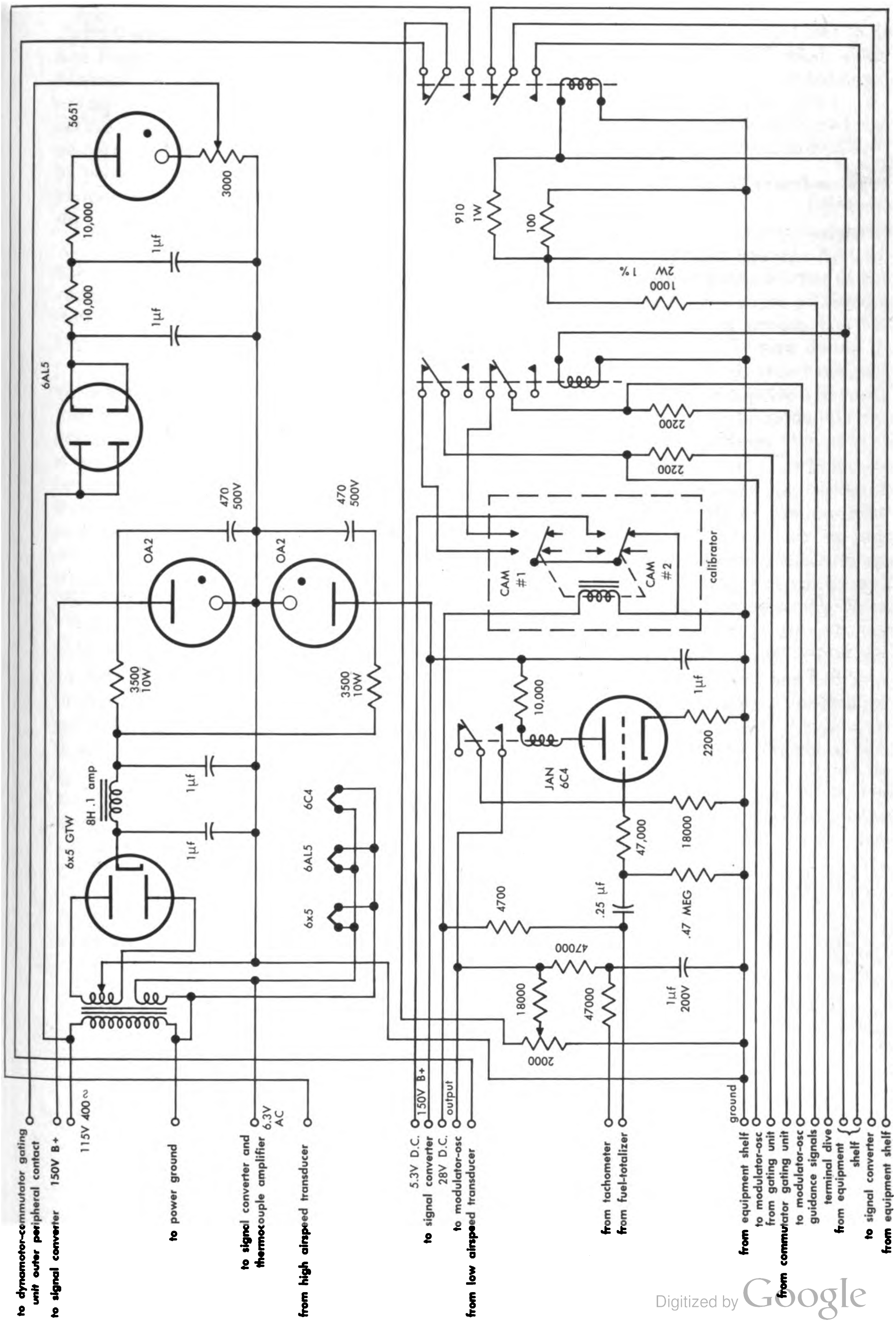
Power required for the telemetering equipment is taken from the P/A's 28-volt DC and 115-volt, 400-cycle lines. A 5.3-volt battery is used to supply reference voltage to the end instruments, and a 24-volt battery is used as a "bucking" battery for telemetering the 28-volt DC power level of the missile's generator.

Power-Supply Unit

The primary purpose of the power-supply unit is to furnish regulated B+voltages of 150-volt amplitude to the 8-channel and single-channel signal converter units, the switching unit, and the thermocouple amplifier unit. The power-supply unit also supplies to these units an output of 6.3 volts AC at 8 amperes. The AC is used for the tube heaters (filaments). The accompanying schematic representation of the power supply unit shows the many outputs as well as other circuits incorporated in the unit.

Also incorporated in the power-supply unit, although not actually power-supply functions, are the terminal-dive switching relays, the cam calibrator, the 115-volt, 400-cycle missile power telemetering reference rectifier, and the fuel totalizer-engine tachometer switching unit.

The 115-volt, 400-cycle input is coupled to the primary of a transformer. The secondary outputs of the transformer are 6.3 volts AC at 8 amperes for heater voltages and a 625-volt, 70-milliampere plate-voltage supply center-tapped to ground. The 312-volt output from each side of the ground-center secondary is coupled to a full-wave rectifier using a 6X5 tube. The cathode DC output is then coupled through a pi-type condenser-input filter con-



Power supply unit of an FM/FM telemetering system

sisting of two 1-μf capacitors and an 8-henry, .1-ampere choke. The voltage is then dropped and regulated for a +150-volt, 24-milliampere and a +150-volt, 12-milliampere output through two 3500-ohm, 10-watt resistors and two OA2 cold-cathode voltage-regulator tubes.

Fuel Totalizer-Engine Tachometer Switching Unit

The engine tachometer AC output and the low- or high-airspeed transducer outputs are coupled to the 5.4-kc modulator-channel mixing circuits. In series with the output of this channel and ground is a sensitive relay and SPDT switch unit. This relay, illustrated in the diagram below, is actuated by a current flow through a 6C4 pulse tube. This pulse tube is normally nonconducting and the relay is normally in *open* position.

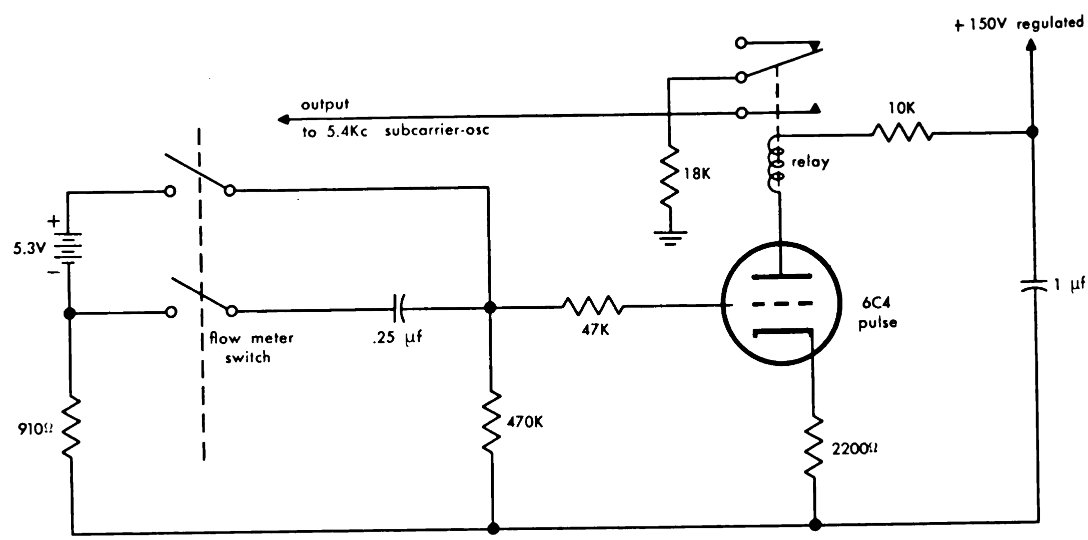
The grid of the pulse tube is coupled to the fuel totalizer unit through a .25-μf capacitor and 47-K resistor. A 470-K resistor from the junction of the .25-μf capacitor and 47-K resistor provides gridleak bias for the tube. A positive voltage is applied to the positive side of the .25-μf coupling capacitor.

The fuel totalizer is a flowmeter-type switching unit which closes a switch every time one gallon of fuel has passed through the unit. In this application, it grounds the positive voltage at the positive (opposite grid) side of the .25-μf coupling condenser through a 910-ohm, 1-watt resistor, thus causing a current to flow through the 470-K gridleak resistor from ground and developing a positive voltage at the grid of the pulse tube. The 6C4 now conducts, causing the series relay in its plate circuit to close. This action lowers the DC output level of the 5.4-kc modulator-oscillator channel for a period of time determined by the RC time constant of the .25-μf coupling capacitor and 470-K gridleak resistor (approximately 0.12 second).

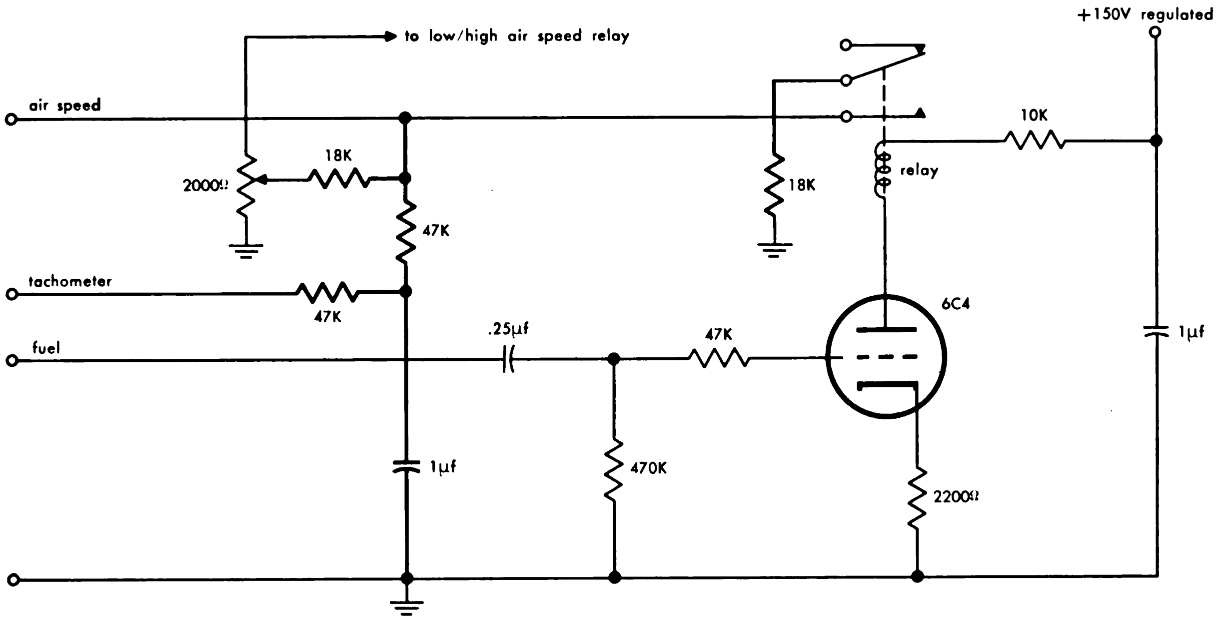
When the information recorded on the telemetering receiving equipment is interpreted, every break in the 5.4-kc channel output designates that a gallon of fuel has passed through the fuel totalizer.

The low- and high-airspeed switching relay is actuated by a 28-volt DC signal from controls at the dump point. Prior to the dump point, telemetering data is obtained from a transducer calibrated for a maximum airspeed of 450 mph. During the terminal dive, airspeed increases; therefore, it is necessary to switch to a transducer calibrated for higher airspeeds. The output of the selected transducer is coupled to the 5.4-kc modulator-channel mixing circuit.

Another function of the airspeed switching relay is to switch the telemetering circuit at the dump point of the azimuth-signal error from the transducer to the accelerometer signal for telemetering data during the termi-



Fuel totalizer switching circuit



Tachometer, low/high airspeed, fuel totalizer mixing circuit (heavy line)

nal dive. The output from the switching relay for this information is coupled directly to one of the eight signal-converter-unit channels.

Tachometer Signal Generator

The AC output of the engine-driven tachometer generator varies in amplitude and frequency from 17 volts, 30 cycles (low engine rpm), to 30 volts, 70 cycles (high engine rpm), the voltage/frequency output being linear to engine rpm. This voltage/frequency output is coupled to the output of the 5.4-kc modulator-channel mixing circuit.

The mixing circuit, shown above, is provided for superimposing the tachometer and fuel-totalizer pulse outputs upon the low/high-airspeed signal of the 5.4-kc modulator channel. The mixer circuit is a resistive network consisting of four resistors and a capacitor. A 47-K resistor and a 1.0-μf capacitor are used as an L-type RC output filter-attenuator for the tachometer. The output of this filter attenuator is fed through a 47-K resistor in parallel with the other two legs of the mixing network.

Low/high-airspeed and fuel-totalizer signals are fed through 18-K resistances. This permits the sine-wave output of the tachometer and the pulse output from the fuel totalizer to be superimposed on the DC level of the low/high-airspeed output.

Cam Calibrator

The calibrator circuit, illustrated in the schematic drawing of the power supply unit on page 509, consists of a motor-driven cam which actuates two microswitches for applying 5-volt and 0-volt signals alternately as flight calibration for telemetering. The motor cams and switches are adjusted so that 5-volt and 0-volt pulse indications appear alternately every 38 ± 8 seconds and remain for 2 ± 0.5 seconds. The cam motor operates on the missile's primary 28-volt DC supply. These *time-calibration* breaks appear on the outputs of the 7.35-kc and 14.5-kc modulator channels.

A relay is actuated by a signal from controls at the dump point, thus removing the cam calibrator from the circuit.

Nozzle-Temperature Thermocouples and Amplifier Unit

The purpose of the nozzle-temperature thermocouple unit is to interpret for telemetering purposes the nozzle-temperature output of the jet engine exhaust.

The thermocouple unit is located at the extreme aft end of the missile. It is attached to a bracket which allows the unit to extend down into the exhaust stream as the stream emerges from the engine blast tube. This unit consists of three thermocouples mounted in a staggered array so as to record the center, intermediate, and outer temperatures of the nozzle output. The three thermocouples comprising this *rake* unit are connected in series, and their voltage output is applied in series with that of four thermocouple units located in the plenum area. These four thermocouples are also connected in series and are used for reference and stabilization. The voltage output of the nozzle *rake* plus that of the four thermistors in the plenum area is then connected to the thermocouple chopper amplifier.

The voltage input to the thermocouple amplifier varies from 45 to 85 millivolts, dependent on the heat output of the engine exhaust. This input voltage is connected to a 400-cycle chopper. From the chopper, the voltage is fed to three amplifier stages. The amplified, 400-cycle voltage is then transformer-coupled to a diode rectifier. The DC output of the rectifier is smoothed and filtered in an RC filter circuit, adjusted by a potentiometer, and coupled directly to the commutator-gate unit. The amplified output which is adjusted to vary between 0 and +5.3 volts, minimum to maximum, is linear to the voltage input from the thermocouple.

Signal Converter Unit

The purpose of the signal converters is to convert sine-wave and square-wave voltage signals into 0- to 5-volt DC signals which maintain a linear relationship to the phase characteristics of the applied AC signals. The AC signals applied to the signal converters originate in guidance and control units.

Signals received from controls are 400-cycle AC signals and represent the following data:

1. Yaw error.
2. Roll error.
3. Pitch error.

Signals received from guidance are 400-cycle AC signals and represent the following data:

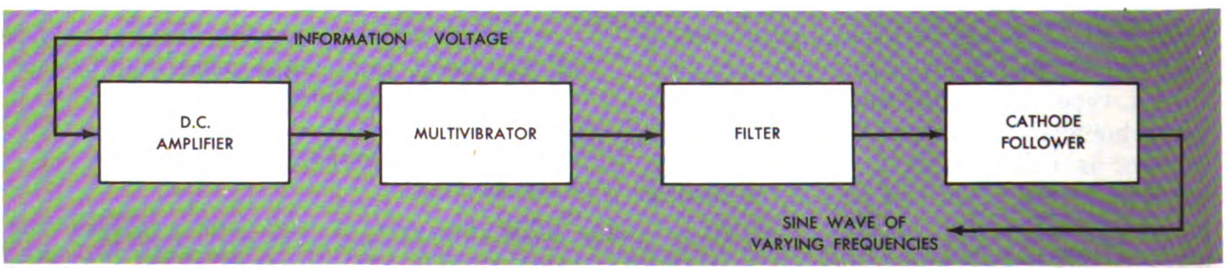
1. Airspeed error.
2. Azimuth error.

The signal converter consists of nine separate identical channels. For circuit analysis and components identification, channel 1 is taken as an example. It is illustrated on the right.

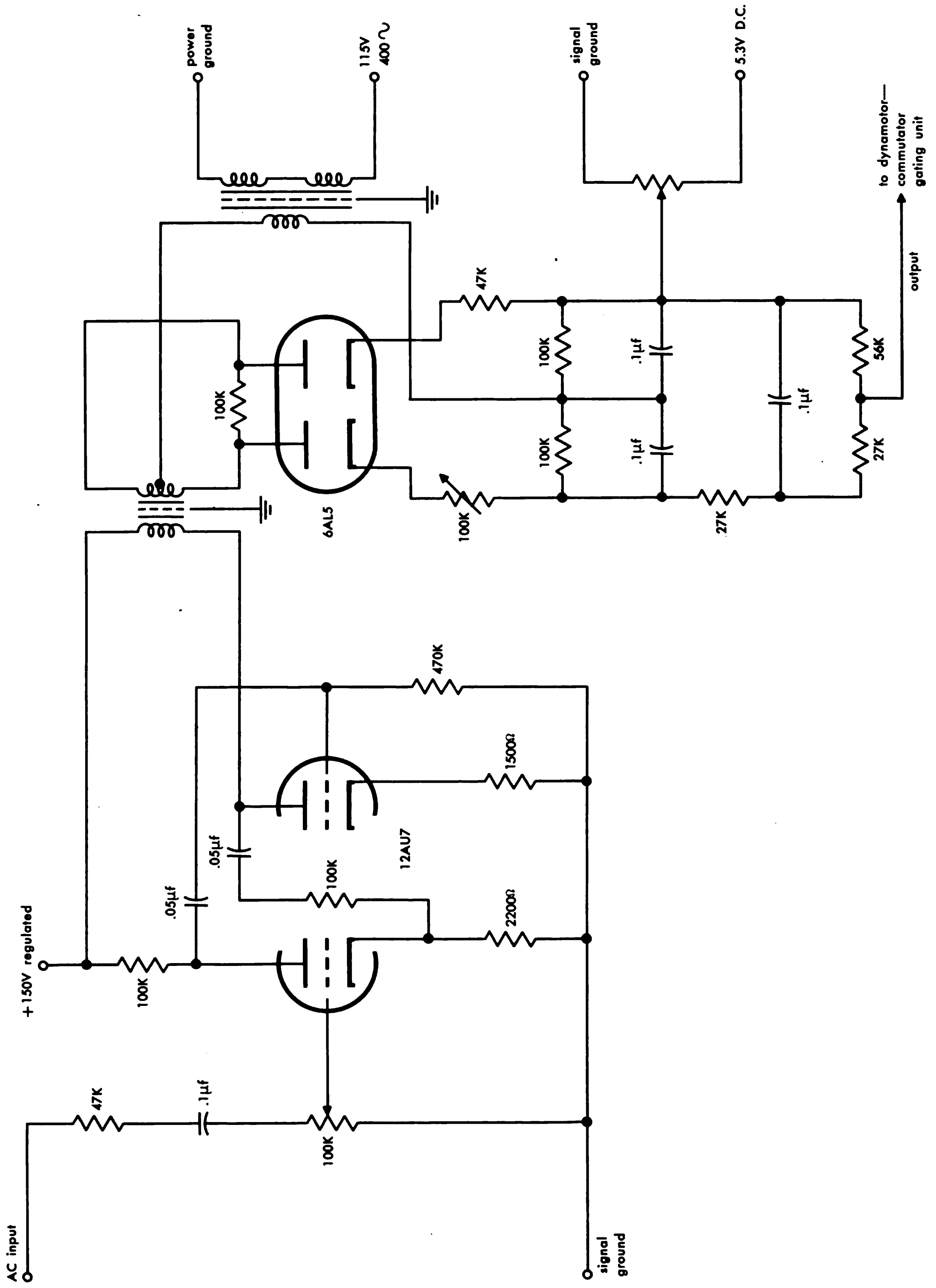
The AC signal input from guidance or control units is coupled through an input amplifier circuit for adjusting input level. Two triode amplifiers with inverse feedback, each being half of a 12AU7 dual triode, comprise the amplifiers. Their output is transformer-coupled to a dual-diode (6AL5) rectifier circuit. The DC output from the cathode of the 6AL5 rectifier is filtered and smoothed in an RC filter network. The output is adjustable for a predetermined DC level for a given phase condition of the input signal.

Voltage-Controlled Subcarrier Oscillators

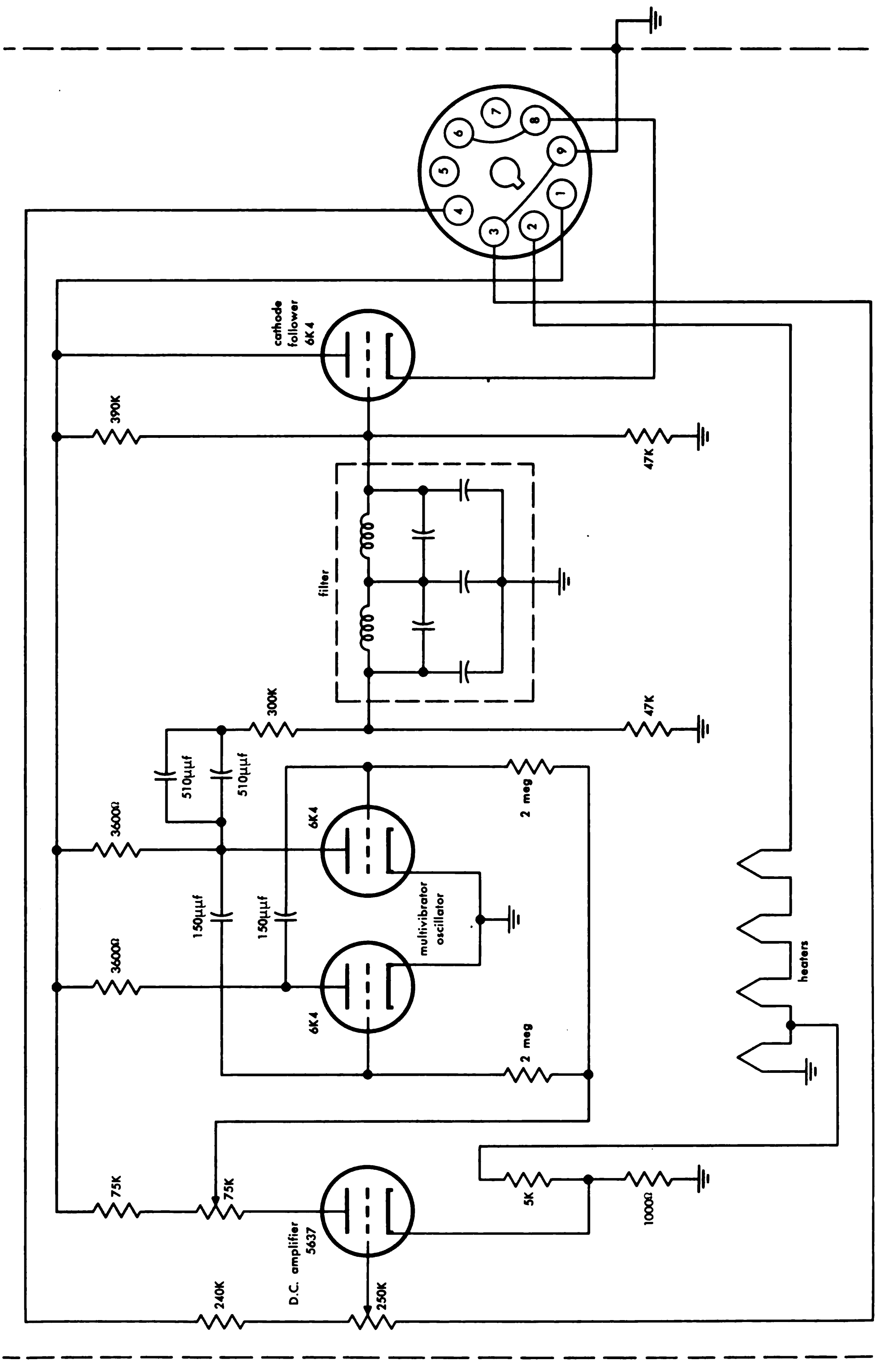
The voltage-controlled subcarrier oscillator are voltage-sensitive oscillators. Their purpose, as indicated by the accompanying block diagram, is to receive information in the form of a varying DC voltage and transform this voltage into a sine-wave AC voltage, varying



Subcarrier oscillator unit



Channel 1 of signal converter



Subcarrier modulator oscillator unit (plug-in type)

in frequency. The frequency variation is dictated by the information voltage.

In other words, the oscillators are frequency-modulated by varying DC voltages from potentiometer-type pickups or commutated information from commutator-gating units. A frequency deviation of $\pm 7\frac{1}{2}\%$ (or 15% total) about the center frequency of the oscillator is produced.

The schematic diagram at the left shows information input voltage from the potentiometer-type pickup. This voltage normally varies from 0 to +5 volts and is applied across a 0.25-megohm potentiometer to the grid of a 5637 triode DC amplifier tube. The output of the DC amplifier tube varies the DC level on the grids of the two 6K4 triode multivibrator tubes shown in the diagram. This causes the frequency of the multivibrator to shift proportionally over the range of the oscillator by $\pm 7\frac{1}{2}\%$ from the center (normal free-running frequency) frequency for an input of 0 to 5 volts.

When voltage is modulated by commutated information in the form of pulses, the higher frequency channels are customarily used. In this manner, higher-frequency responses are obtainable. With the pulses from the commutator-gating units varying from −0.4 volts to +5.3 volts, a maximum deviation of $-7\frac{1}{2}\%$ to $+7\frac{1}{2}\%$ of the center frequency is obtained.

The square-wave signal from the multivibrator is passed through a low-pass RC-filter network to remove harmonics, so as to produce a sinusoidal output with less than 0.6% total harmonic distortion. A cathode-follower stage employing a 6K4 triode provides a low-impedance output.

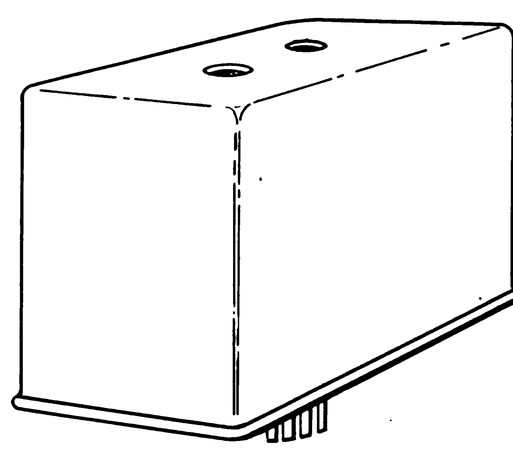
These oscillator units employ a filament voltage-compensation network in the cathode circuit of the DC-input, amplifier-type 5637 triode. This arrangement restricts center-frequency drift of the oscillator to a small percentage of the bandwidth for a drop of as much as 20% in the filament (heater) supply voltage. The compensation is accomplished by returning the cathode to a voltage supply. As the heater voltage drops and emission from the cathode falls off, the drop in heater current through the cathode self-bias resistors reduces the cathode potential. As a result, the cathode is made less positive with respect to the grid, producing the same effect as if the grid had been made more positive with respect to cathode.

This effectively positive signal on the grid keeps the plate current constant regardless of changes in cathode temperature, and it reduces the effect of drop in heater voltage upon tube amplification.

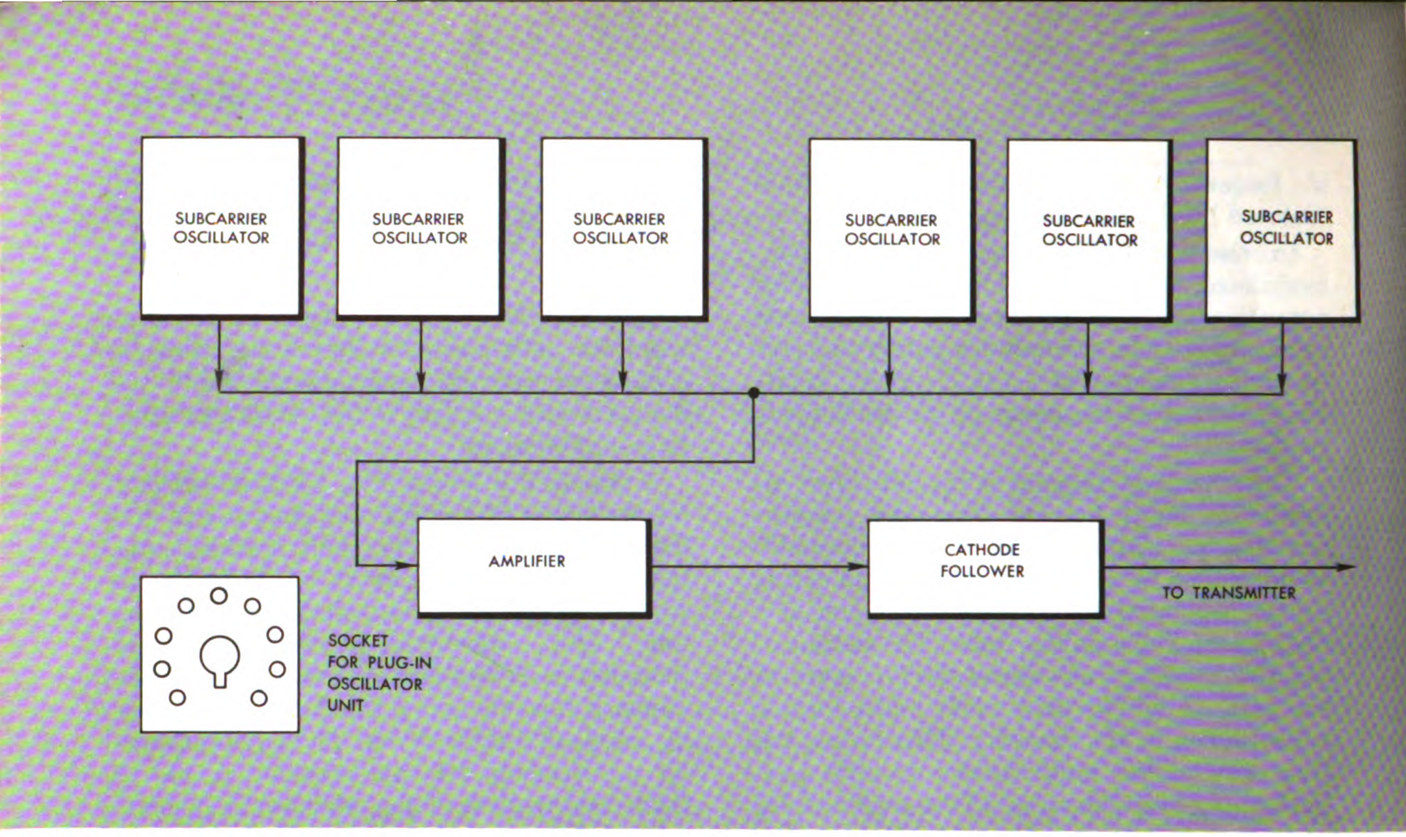
A standard nine-pin plug, pictured directly below, which is located and centered in the bottom of each oscillator-unit shielded container, permits the units to be plugged into sockets on the subcarrier-oscillator mounting unit. Two adjustable potentiometers are mounted in each oscillator unit. They can be adjusted from the top of the shielded container by means of a screwdriver. These potentiometer adjustments set the deviation points or range of deviation of the oscillator frequency.

The subcarrier oscillators are identical in circuitry. They vary only with respect to the component values employed to obtain the desired channel or frequency.

Variations and modifications of the subcarrier oscillators have been tried in an attempt to improve their stability to the highest degree. The type described here is most widely used at the present time. Next to the commutator, subcarrier-oscillator units have been the most critical components in telemetering systems.



Oscillator container



Subcarrier-oscillator mounting unit

Subcarrier-Oscillator Mounting Unit

The block diagram above shows the oscillators mounted on the subcarrier-oscillator mounting unit. The unit has two main functions in addition to serving as a mounting base: to supply input power and connections for six subcarrier oscillators, and to mix and amplify the FM output of the oscillators to frequency modulate the RF transmitter. The schematic drawing on the following page illustrates interconnections of these oscillators.

The mounting unit provides bakelite plates for firm seating of the oscillator units. It also provides a metal cover plate for clamping down the units securely to reduce the effects of vibration and acceleration.

Six nine-pin sockets are provided. The oscillator units are plugged into these sockets. An external cable from pin 4 of each socket provides for connecting each oscillator unit to its end instrument or data pickup.

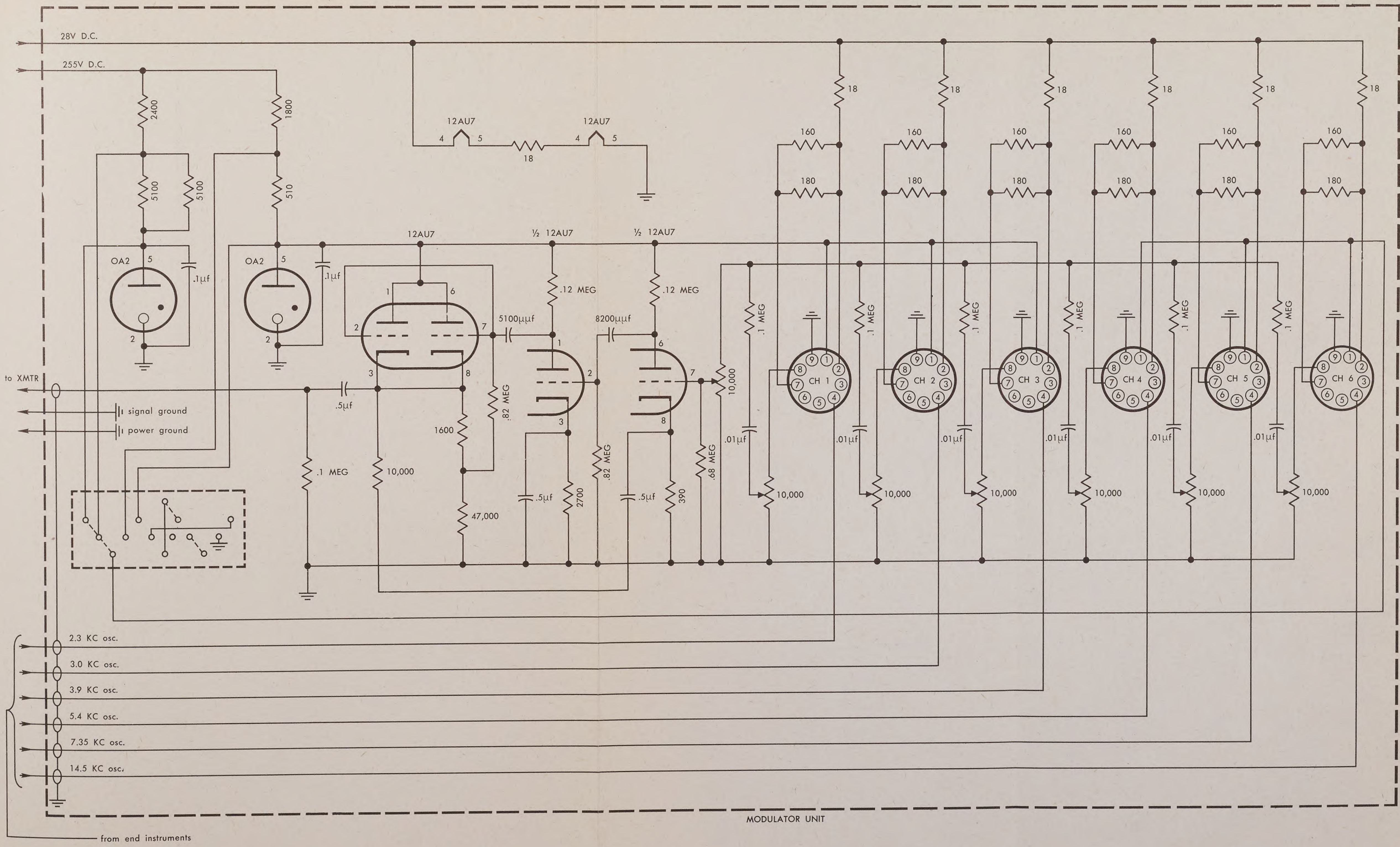
As mentioned previously, the information from the pickup modulates the frequency of the oscillator. The schematic of the mounting unit shows how the output signals from pin 8 of all of the oscillator units are fed through individual output potentiometers of 10,000 ohms each. The potentiometers are used for *preemphasis* adjustment. From here, the signal passes through RC decoupling networks to one 10,000-ohm common-output potentiometer for mixing. This common-output potentiometer acts as the *master gain control* for amplifier stages of the mounting unit and as the *deviation control* for the RF transmitter.

The mixed frequency-modulated subcarrier signals are amplified in two amplifier stages, utilizing a type 12AU7 dualtriode. A cathode-follower output stage, using a 12AU7 tube with its two triode sections connected in parallel, provides a low-impedance output to the transmitter.

Negative feedback from the cathode-follower output stage to the first amplifier stage through a 10-K resistor and a 0.5-μf capacitor reduces the harmonic distortion to less than 0.5%.

The frequency response curve of the amplifier is flat within one db up to 85 kc.

Plate voltage for the amplifier and the subcarrier oscillator is obtained from the +225-volt supply from the dynamotor-commutator-gating unit. The + 225-volt supply is dropped


28V D.C.
255V D.C.
2400
1800
5100
5100
510
OA2
OA2
.1μf
.1μf
12AU7
½ 12AU7
½ 12AU7
.12 MEG
.12 MEG
5100μμf
8200μμf
.82 MEG
.82 MEG
.68 MEG
1600
10,000
47,000
.1 MEG
.5μf
.5μf
.5μf
2700
390
to XMTR
signal ground
power ground
18
160
180
.1 MEG
.01μf
10,000
CH 1
CH 2
CH 3
CH 4
CH 5
CH 6
2.3 KC osc.
3.0 KC osc.
3.9 KC osc.
5.4 KC osc.
7.35 KC osc.
14.5 KC osc.
MODULATOR UNIT
from end instruments

: -150-volts by a re
gulated by two OA2

Filament voltage i
volt DC line. The osc
to provide both 12.6
for the different types
controlled, strain-ga
ance – as required.

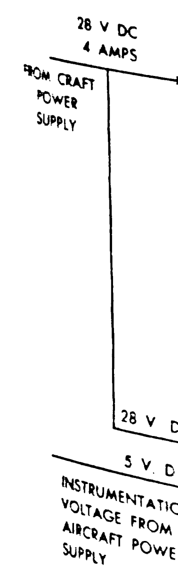
A 5.0-volt instrum
ery for the poten
could be supplied
but if necessary, it c
by the mounting uni

Dynamotor-Commuta

The dynamotor-
shown in block diag
of three sections:

1. A dynamotor
voltage for all the
amplifier in the st
mitting system.

2. A commutat
synchronizing mas
pulses for the gating
to switch the 5.0-vo



to +150-volts by a resistor network and then regulated by two OA2 voltage-regulator tubes.

Filament voltage is supplied from the 28-volt DC line. The oscillator sockets are wired to provide both 12.6 volts and 25.2 volts DC for the different types of oscillators — voltage-controlled, strain-gauge, and variable-reactance — as required.

A 5.0-volt instrumentation voltage is necessary for the potentiometer pickoffs. This should be supplied by an external battery; but if necessary, it can be supplied internally by the mounting unit.

Dynamotor-Commutator-Gating Unit

The dynamotor-commutator-gating unit, shown in block diagram form below, consists of three sections:

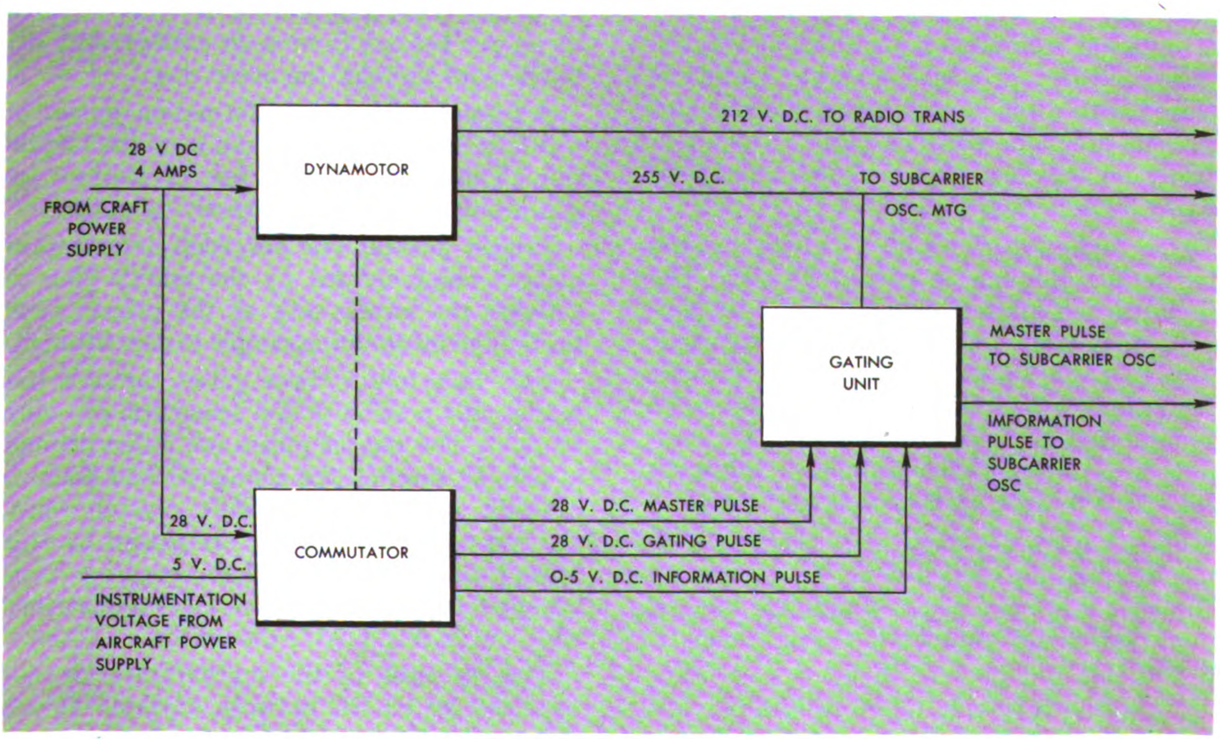
1. A dynamotor section to supply B + voltage for all the units except the final RF amplifier in the standard 32-channel transmitting system.

2. A commutator section to provide a synchronizing master pulse and to trigger pulses for the gating section. Also, it is needed to switch the 5.0-volt instrumentation voltage to each of 27 pickoffs separately and in succession. It collects an information sample from each pick-off to feed the gating unit at the rate of 5 samples per pickoff per second.

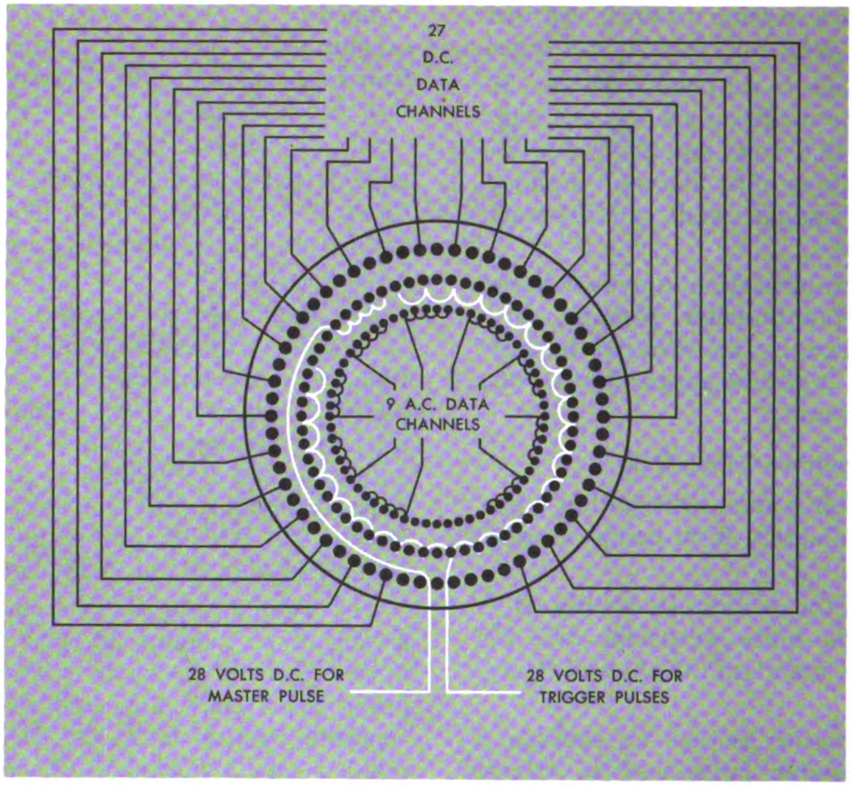
3. A gating section to pass the sampled information pulses, electronically maintaining them at a constant width. These pulses are applied to the integrating circuits in a receiving system.

The dynamotor converts 4 amperes at 28 volts DC to 235 milliamperes at 355 volts DC. Included in the unit, as part of the dynamotor section, are two filters. One is an input filter that reduces conducted RF line noise and prevents any RF noise from the commutator and brushes from appearing on the 28-volt DC line. The other filter reduces the ripple in the output to less than 0.5%.

The commutator is a circular plate consisting of six concentric rings made of coin silver. Three of these are solid rings, and the other three rings are each made up of 60 separate circular contacts. All six rings are molded into a flat plate of *Mycalex 410* or some other suitable insulating material. The schematic diagram on the following page



Dynamotor-commutator-gating unit



Telemetering commutator

illustrates a telemetering commutator with its internal contacts and interconnections. Each solid ring and each separate contact is backed with a pin to which the external connections are soldered. The plate is mounted rigidly on the head of the dynamotor, positioned by a locating pin and slot.

The commutator rotor is driven by the dynamotor through a reduction gear. The rotor has self-lubricating silver-graphite contacts mounted on phosphor-bronze wiper arms to maintain good contact and provide clean samples. Very low voltage signals can be commutated satisfactorily, since the commutator is unusually free from noise and small thermal and brush contact voltages.

Three separate circuits, one through each pair of rings, are switched at the same time. These rings are grouped as inner, middle, and outer peripheral contacts in the schematic above.

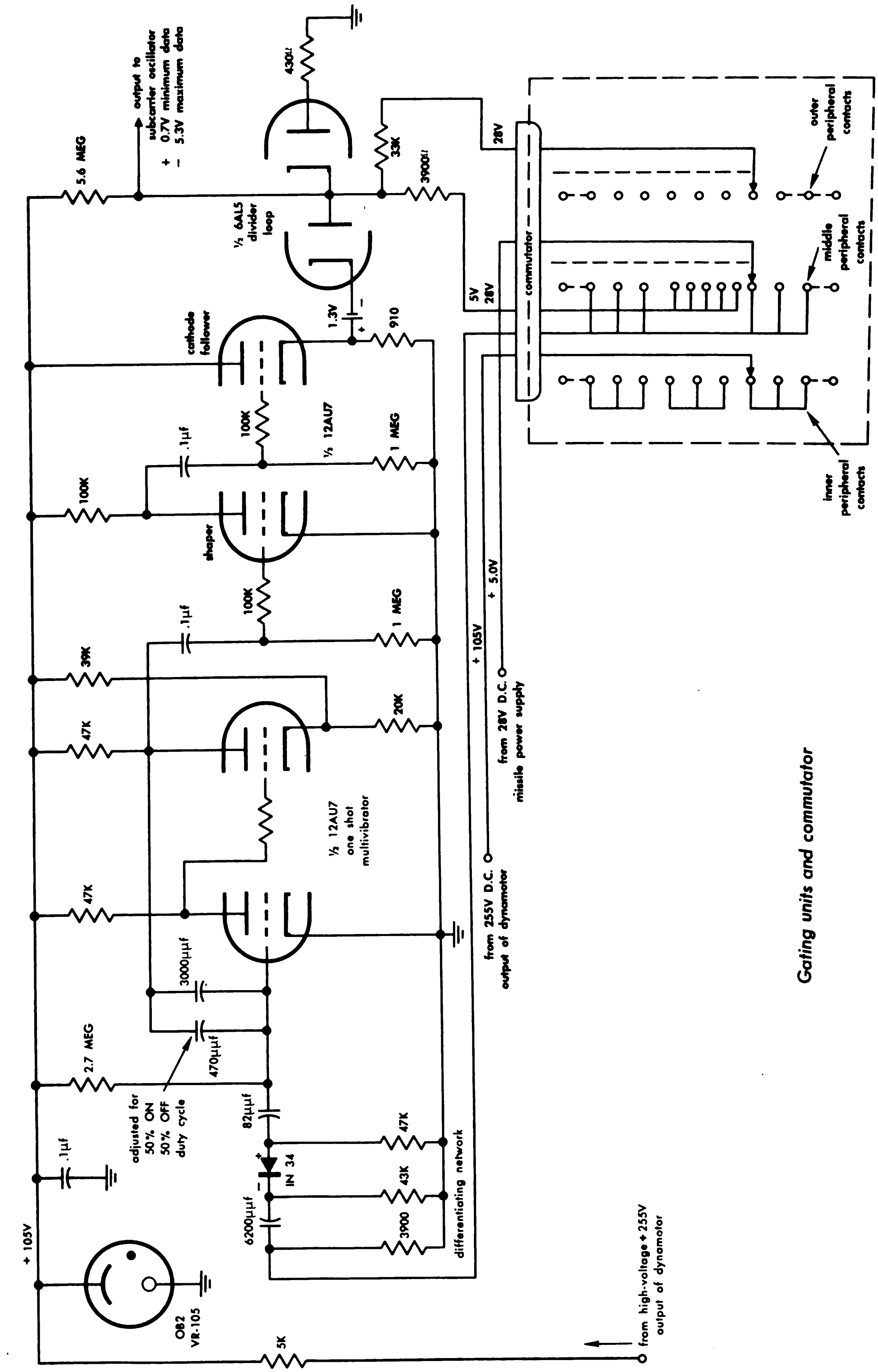
A positive five-volt instrumentation voltage is supplied from the inner ring to each commutated pickoff in succession for sampling purposes. The information is sampled five times per second from each pickoff and fed to the gate through the outer ring.

A master pulse, five information pulses in width, is fed to the gating unit from the middle ring. This is done once per revolution for synchronization with the receiving station.

Trigger pulses, one for each information pickoff, are generated by the middle ring to activate the gating unit at the proper time to allow an information sample to pass through.

Now let's follow these pulses through the gating-unit portion of the schematic drawing at right. The trigger pulses are differentiated in passing through the condenser and resistor network preceding the IN34 crystal. Only the *negative portions* (pips) are passed by this crystal diode rectifier. These negative pips are used to turn on the *one-shot multivibrator* which utilizes a 12AU7 dualtriode.

The ON time of the multivibrator is determined by an RC network adjusted at the factory for 50% ON time, 50% OFF time, at 6000 rpm. The negative pulse from the



Gating units and commutator

multivibrator is then fed through a *shaper stage* (one section of a 12AU7 dualtriode) for squaring the pulses. From there, the signal goes to the cathode follower (second section of the 12AU7) which is regularly cut off for half of each cycle by negative pulses from the shaper stage.

When the cathode follower is cut off, a small current flows around a divider loop formed by two diodes (6AL5 dual diode) and two resistors (one 430 ohms and one 910 ohms)with a 1.3-volt battery as the voltage source.

Approximately −0.4 volt appears at the junction of the diodes between pulses. This is the level between pulses and is called the *reference level*.

Since the diodes, when conducting, present a low impedance to ground, any positive information appearing at the junction of the diodes is bypassed to ground through one section of the diode.

When the cathode follower conducts due to a positive gating pulse from the preceding stage, the diodes do not conduct, and they present a high impedance to the ground. The information pulses of 0 to +50 volts from the outer ring of the commutator appearing at the junction of the diodes, are passed to the voltage-controlled subcarrier oscillator, which is fed by the output of this unit.

Minimum information, also known as the *blanking level*, then actually appears at the diode junction as +0.7 volt. This level appears as +0.7 volt due to the voltage drop between B+ (105V) through the 5.6-megohm dropping resistor and 33,000-ohm resistor to ground. The level is approximately 20% of the total pulse height. A blanking level is required for synchronization between the transmitting and receiving equipment, and it serves as a pedestal signal upon which the date pulses are superimposed.

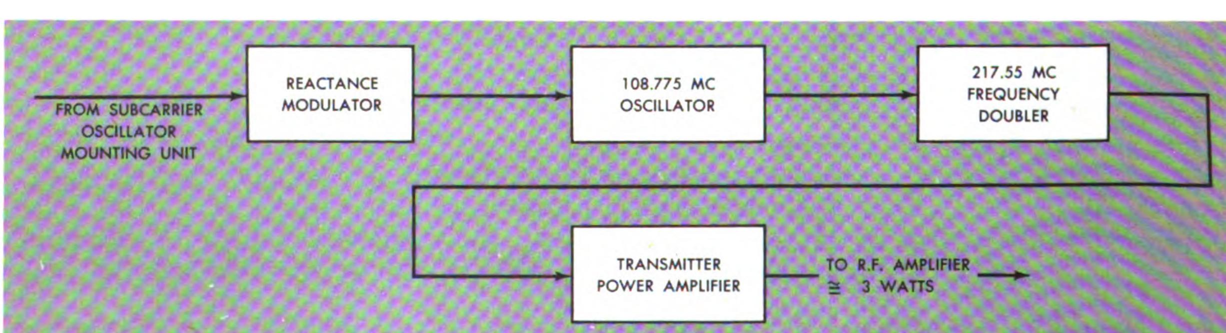
Maximum information level desired appears as +5.3 volts at the junction of the diodes. This is due to the effect of the 5.6-megohm and 33,000-ohm resistors, the 490,000 input impedance of the voltage-controlled subcarrier oscillator, the 105-volt B+, and 5.0 volts from the commutated pickoff. The arrangement of the divider system is designed to produce this desired output voltage.

Only the *center portion* of any information sample is allowed to pass through the diodes; thus, any noise which might be present on the leading or trailing edges of the information pulse is eliminated.

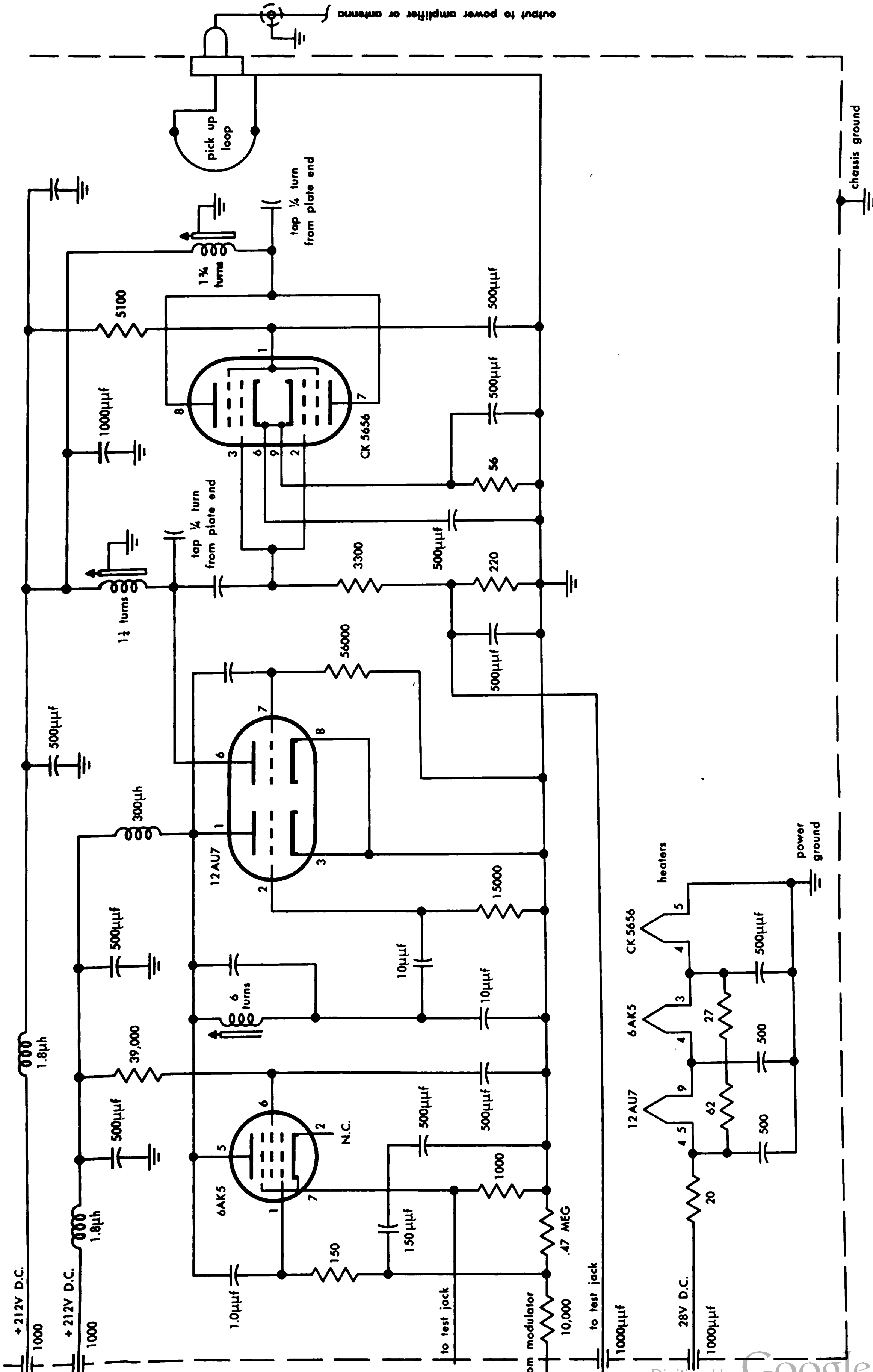
Since the ON time of the multivibrator can be accurately adjusted, the time of conduction of the diodes is accurately determined. Therefore, the gating unit passes only pulses of a constant width, even with a drop in commutator speed or a drop of as much as 20% in the supply voltage.

The height of each pulse is then directly proportional to the information developed across the individual data pickoffs.

The master pulse from the middle ring of the commutator appears at the gate when it is closed (i.e., the cathode-follower cutoff), but passes through since it is originally 28 volts in peak amplitude and drops across a smaller resistor (3000 ohms) than the information pulses do. It overcomes the negative



Transmitter unit



Telemetering transmitter unit (3 watts)

voltage at the junction of the diodes and appears at the output of the gating unit, although at a considerably reduced voltage.

An OB2 (VR-105) voltage-regulator tube maintains a constant B+ supply of 105 volts.

Transmitter Unit

The transmitter utilizes the output of frequency-modulated subcarrier oscillators to frequency modulate an RF signal produced by the transmitter. The modulated signal is then doubled (or multiplied) in frequency and amplified for transmission. The center frequency of the resultant RF carrier is adjustable within the range of 215 to 230 megacycles.

The mixed frequency-modulated subcarriers are received from the subcarrier-oscillator mounting unit. They are fed either to the grid or to the cathode of a reactance-tube modulator, which is the 6AK5 in the schematic diagram on the preceding page. The 6AK5 is a miniature pentode.

Grid input is used due to its high impedance, which matches the output of the amplifier in the subcarrier-oscillator mounting unit.

Cathode input to the transmitter is provided so that the unit can be used with a low-impedance modulator if desired.

As the signal varies in amplitude, the reactance-tube modulator (6AK5) acts as a varying capacitance across the tuned circuit of a modified Colpitts oscillator employing one section of a 12AU7 dual triode. This varying capacitance shifts the frequency of the oscillator in accordance with variations in amplitude of the input signal.

Output of the oscillator is fed to a frequency-doubler stage employing the second section of the 12AU7 dual triode. Output of the frequency doubler drives a power-amplifier stage. The power amplifier uses a type CK5656 dual tetrode with its two sections tied in parallel. The dual tetrode maintains a power output level of at least 3 watts over the entire frequency range when it is fed into a 51.5-ohm load through an inductive output-coupling link.

The output of the 3-watt power-amplifier stage can be fed directly into a 51.5-ohm load such as a quarter-wave antenna, or it can be used to drive an RF amplifier of higher wattage output.

The tuned circuits of the oscillator, doubler, and power amplifier are all adjustable with permeability (slug) tuned coils.

A filter box surrounding the cable that contains all inputs and power leads reduces the possibility of RF leaking back to the other units of the system. Noise in the transmitter amounts to about 5 kc of the +125-kc deviation at approximately 10 G's under vibration of 10 to 60 cycles per second, but it occurs mainly at frequencies below the lowest subcarrier frequency employed.

RF Power-Amplifier Unit (35-40 Watt)

The RF amplifier is used to increase the power output of 3 watts from the transmitter unit to 25 watts, over a frequency range of 215 to 230 megacycles.

In the schematic drawing at the right of an RF power-amplifier unit, the output of the transmitter unit drives an 829-B dual pentode operated in push-pull. The output of the 829-B is applied to the antenna through the coupling link matching the 51.5-ohm output impedance. The coupling link can be tuned to the transmitter frequency by a variable capacitor of 2 to 4 $\mu\mu f$ capacitance.

The 829-B amplifier tube is operated with 500 volts DC on the plate from its own dynamotor power unit. The plate circuit can be tuned for maximum power output at the desired center frequency by means of a variable inductor. The amplifier is designed to operate into a 51.5-ohm load.

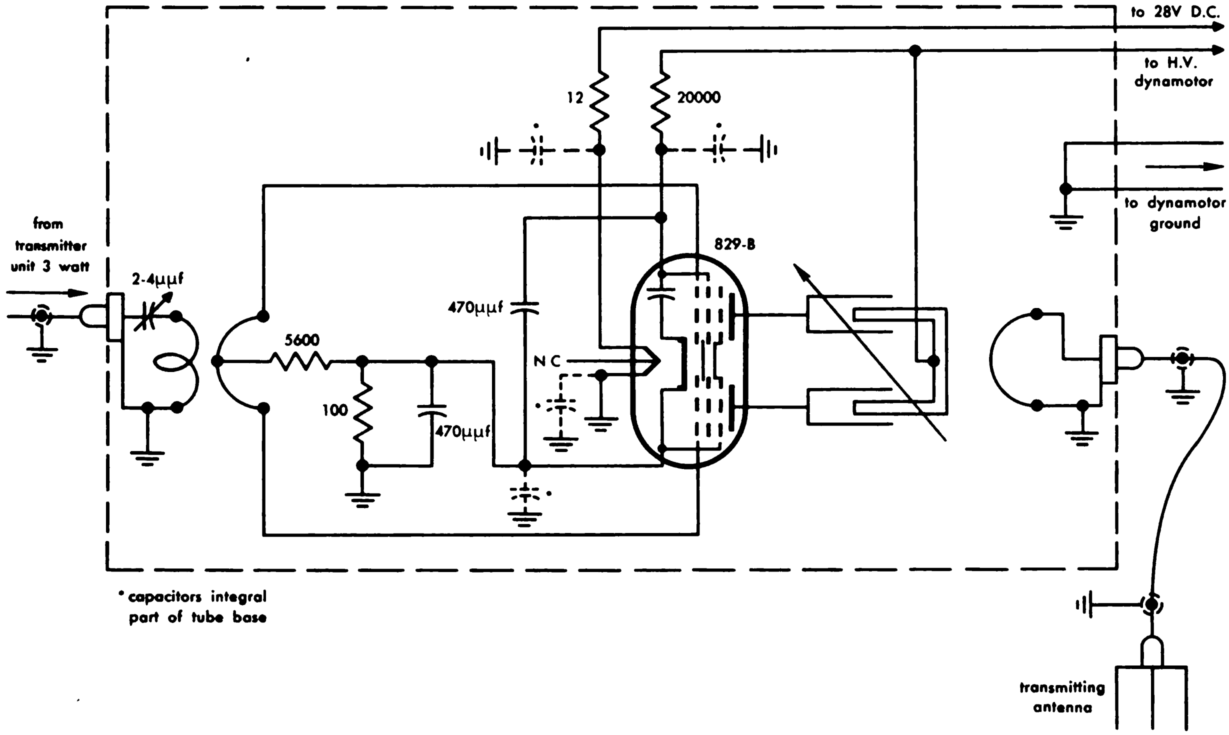
Bypass condensers from the cathode and heater-pin connections to ground reduce the possibility of regeneration as well as the presence of RF on the filament lines.

Dynamotor

The high-voltage dynamotor converts 6 amperes at 28 volts DC to 200 milliamperes at 500 volts DC. An input filter network helps reduce conducted RF line noise and prevents RF line noise from appearing on the 28-volt line. An output filter network restricts the ripple to less than 0.5%.

Switching Unit

An electronic switching unit is provided to indicate to the telemetering system the time at which the dump signal is initiated. This is



RF power amplifier unit (35-40 watts)

accomplished by switching the 2.3-kc subcarrier modulator channel from the yaw-error pickup to a +5.3-volt DC indication.

A 28-volt DC voltage is coupled to the switching tube (6AU6) through a 1-μf capacitor and a 470-K resistor. This voltage originates from the missile control system. At the time of application of the 28-volt signal, the 1-μf capacitor charges. It places a positive voltage on the grid of the 6AU6. This causes the tube to conduct and thus actuate a relay in its plate circuit which applies the +5.3-volt signal to the 2.3-kc channel. The tube conducts for the RC time constant (about 0.5 second) of the 1-μf capacitor and the 470-K resistor to ground.

Upon resuming the stable, nonconducting condition, yaw-error indication continues on the 2.3-kc channel.

Antenna

The antenna generally employed with an FM/FM telemetering system is a ¼-wave stub antenna which is 13½ inches long. It is suitably mounted beneath the missile to radiate power in an unobstructed path to the ground station receiver.

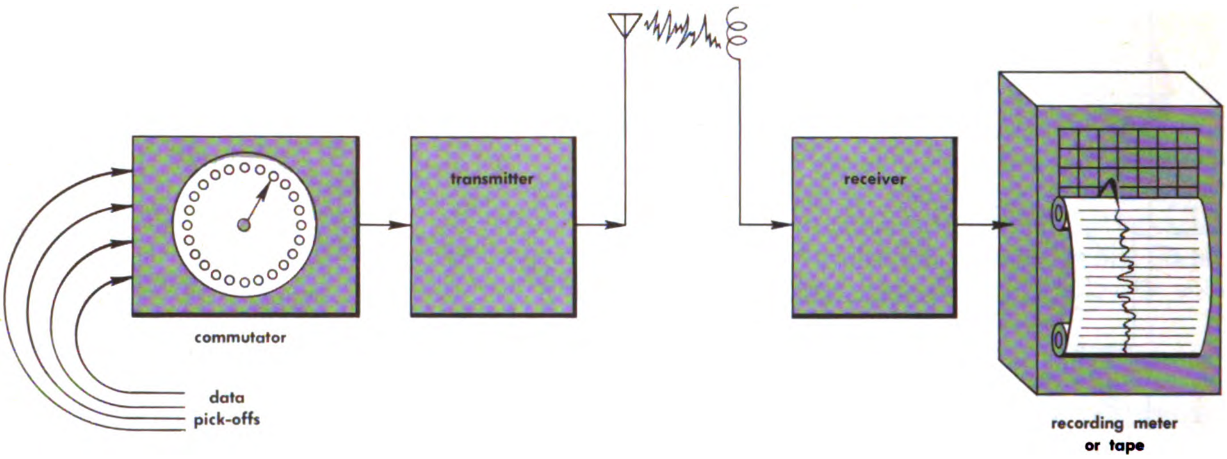
This rounds out our coverage of the airborne components of the FM/FM telemetering system, the system most commonly used in missiles. Let's now go on to another type of telemetering system.

AMPLITUDE-MODULATED (AM) TELEMETERING SYSTEMS

Amplitude-modulated (AM) RF-carrier telemetering systems are used in some industrial applications in which the data requirements are not critical and in which the information samples can be averaged over relatively long periods of time. By averaging over long periods of time, the effects of noise and interference can be minimized.

Such data as water levels in reservoirs, fuel levels in storage systems, YES and NO data indicating action or lack of action, and opening and closing of electrical circuits can be telemetered by AM systems.

However, in telemetering data from research missiles, many samples of data of widely differing characteristics often must be taken in a brief period of time and under widely differing conditions of temperature, pressure, acceleration,



Functional diagram of an AM telemetering system

etc. This data must be telemetered with the highest degree of accuracy. Too, the sampling time is usually short, and physical conditions are complex. Therefore, to insure the greatest accuracy in telemetering such data, the FM/FM systems or pulsed radar systems have proven more suitable than AM systems.

The intelligence conveyed by the RF carrier in an AM system is represented by the amplitude of the carrier. Therefore, any change in carrier amplitude produced by static or from sources outside the system might be mistaken for changes taking place at the data-pickoff points. This results in erroneous interpretations. These conditions preclude the use of AM telemetering systems in missile research.

Circuitry of AM telemetering systems follows conventional lines and resembles that of the usual AM radio-communication systems. Some synchronized switching and timing system must be employed so that the amplitude variations of the received signal at any given instant may be identified with the channel or data source being sampled at that particular time. For this purpose, a commutating system similar to that used in an FM/FM system is employed.

Received data is recorded on tape or charts by recording meters such as the Esterline-Angus meters, which employ a time-division roll chart. A functional block diagram of an AM telemetering system appears above.

PHASE-MODULATED (PM) TELEMETERING SYSTEMS

Phase-modulated (PM) telemetering systems are similar in operation to FM/FM systems; and the same types of transmitting, receiving, and recording equipment can be used.

The PM subcarrier oscillators are more suitable for signal inputs from high-impedance data pickoffs.

While in FM transmission the deviation is the same for all audio frequencies of the same amplitude, in PM transmission the deviation is proportional to the audio frequency. If a 12,000-cycle signal input to the PM system results in a certain frequency deviation, a 3000-cycle signal of the same amplitude would give only one-fourth as much deviation. A 3000-cycle signal of twice the amplitude would still give one-fourth as much deviation. Thus, in a PM system, each subcarrier channel always produces a frequency deviation in the carrier of a nearly constant value. Thus, a channel can be identified by this constant deviation at the receiver without the need for a timing and synchronizing link.

The drawback of the PM system should now be apparent to you; variations in the amplitude of DC pickoff data do not produce a variation in the carrier frequency, and the degree of the data variation can not be de-

termined. PM is satisfactory for direct pick-offs from AC circuits or from end instruments supplying data of *varying frequencies.*

Phase modulation is applied to an amplifier stage, while frequency modulation is applied directly to the oscillator. Therefore, PM is readily adaptable to transmitters employing highly stable oscillators, such as those which are crystal-controlled. These oscillators are desirable because of the minimum amount of adjustment required.

However, the amount of phase shift that can be obtained with good *linearity* is quite limited, and the *maximum practical* modulation index is 0.5 at the radio frequency at which the modulation occurs. Because the phase shift is proportional to the modulation frequency, this index can be used only for the highest frequency present in the modulating signal, assuming that at some time all frequencies will have equal amplitudes. Thus, if the output of the PM transmitter is set for a certain *range of deviation* at a high subcarrier channel frequency, the range of deviation may be less than that desired for lower channel frequencies. If the modulation index is set to full permissible deviation of the carrier for the lowest subcarrier channel, the higher frequency subcarriers may produce *sideband splatter* of the carrier, making reception difficult and generating interference. Therefore, it is necessary to change PM signals to FM signals in which the modulation index decreases in inverse proportion to the modulating frequency, or in which the output voltage is inversely proportional to the modulating frequency.

The additional circuitry required and the compensation of decoding and recording devices makes a PM system less suitable for missile telemetering than the FM/FM system. For these reasons, and because of the exacting requirements of a PM system in terms of calibration and stability of power supply voltages, PM systems are used less in telemetering than FM/FM systems.

Also, in missile research, the data variations rarely occur at frequencies higher than 100 cps but may vary widely between 0 to 100 cps. The carrier deviations produced for variations from 0 to 100 cycles from a single data pickoff would be received much as signals from 100 separate channels, and the amount of information for each cycle of change would be negligible.

RADAR TELEMETERING SYSTEMS

Radar telemetering systems are fundamentally pulse systems based on time division and characterized by pulses of high amplitude and short duration (0.5 to 2.0 microseconds).

The radar frequency spectrum provides bandwidth for many channels of data. The pulse-recurrence frequency (PRF) and pulse width determine the number of data samples which may be taken in a unit of time.

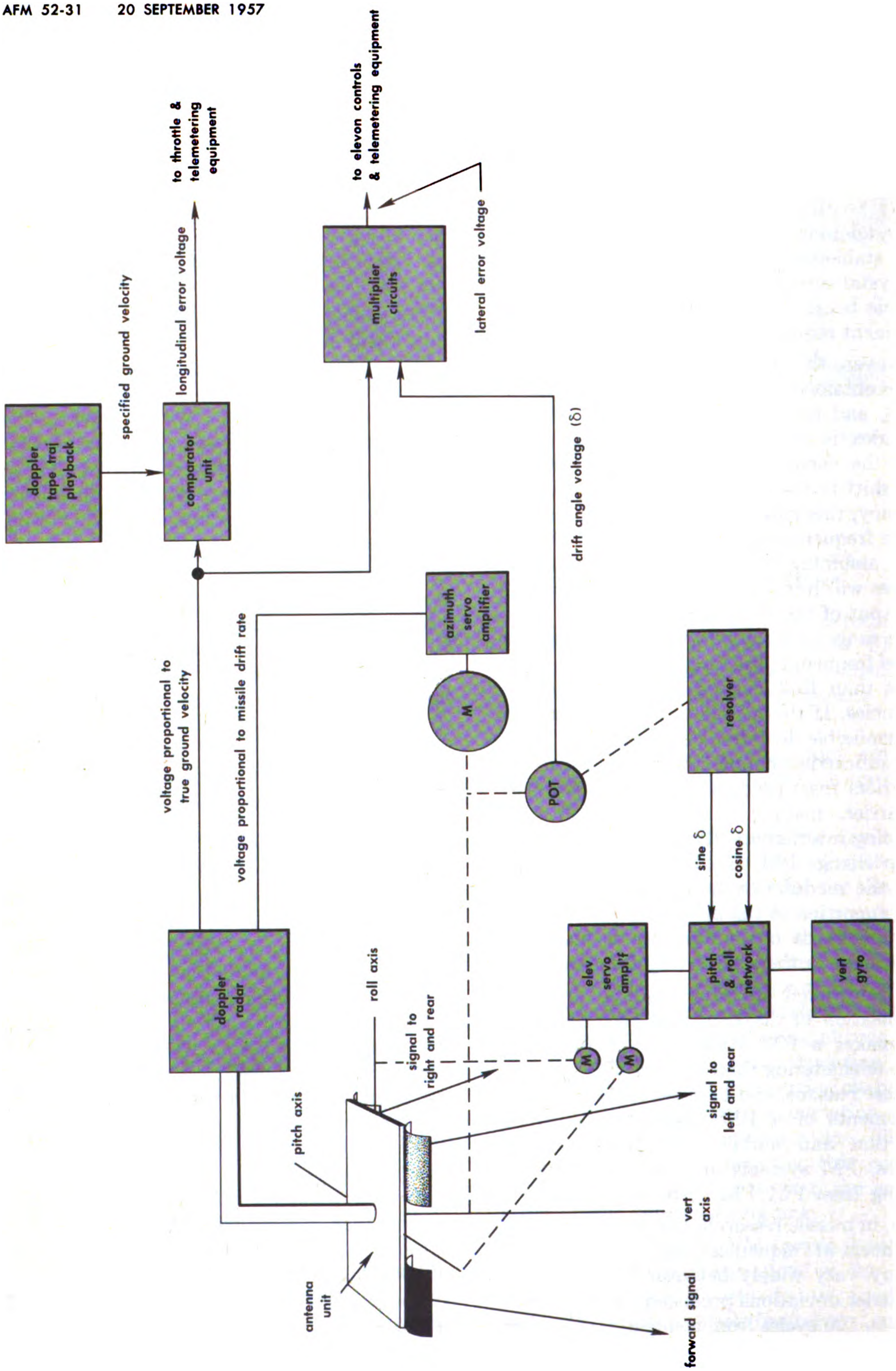
Objections to radar telemetering in missiles research are the high voltages required and the critical nature of components and adjustments with respect to temperature, pressure, timing, and directivity.

Also, the duration of each sampling period is so brief (0.5 to 2.0 microseconds) that only a minute amount of information can be picked off during a single pulse period. However, the frequency spectrum permits many samples from many channels per second of time, so this objection is minimized. For some types of data, the brief sampling period is actually advantageous, particularly in sampling data that is subject to rapid changes.

Pulse-Position Modulation

Radar telemetering systems employing pulse-position modulation offer the most encouragement for missile research. In such a system, each channel is sampled once during each period. The channel pulses are advanced or retarded about their unmodulated positions due to the effect of input data voltages upon the time-base circuits or keyers. The amount of advancement or retardation depends upon the amplitude and polarity of the data voltage and the instant each sample is taken. The input frequency determines the rate at which the pulses corresponding to a given channel vary from their unmodulated position. The maximum change in pulse position must be limited to prevent adjacent channels from overlapping and to reduce crosstalk.

If each channel occupies ten degrees, a total of thirty-five channels, plus one synchronizing channel, can be employed.



Doppler radar system designed for use in missiles

With a pulse width of 0.5 microsecond and a sampling interval of twenty times this amount (10 microseconds), the time per channel is 10 microseconds. Thus, a maximum sampling rate of 2777 samplings per second is obtained (1,000,000 microseconds divided by the number of channels times the time per channel or $\frac{1{,}000{,}000}{36 \times 10}$). A greater pulse width would result in a lower rate of sampling.

Many methods of producing pulse-position modulation are possible. The method used should be precise so that random errors of pulse positioning will be at a minimum. The principal difficulty in pulse-position systems has been in devising absolutely accurate timing circuits and maintaining positive synchronization between the transmitter and ground-station receiving equipment.

Pulsed-Doppler Modulation

The Doppler principle discussed in Chapter 4, Section IV, is employed in a radar telemetering system to accurately determine instantaneous values of range, altitude, and velocity.

A typical Doppler radar system is shown in the diagram on the preceding page. This system is designed for use in a missile. It detects the change in frequency between a transmitted signal and the echo. This frequency change is converted into voltages proportional to groundspeed.

Three antennas are employed, mounted rigidly as one unit. One antenna transmits energy downward and forward; another transmits downward and to the left rear; and the third transmits downward and to the right rear.

The drift voltage is fed to an azimuth servo amplifier which drives a motor to keep the antennas in the same azimuth position in reference to the earth. A potentiometer attached to the airframe, with its arm attached to the antenna unit, supplies a voltage proportional to the drift of the missile. In a multiplier circuit, the true ground-velocity voltage is multiplied by the cosine of the drift angle (δ) to obtain a voltage proportional to lateral error.

External to the Doppler system but essential to its use in the missile is a comparator unit. This unit compares a specified ground velocity from information stored on a magnetic tape with the true ground velocity from the radar.

The output of this comparison circuit is a voltage proportional to longitudinal error voltage. Lateral and longitudinal error voltages, fed to the control system, change the throttle and pitch control positions in such a manner as to eliminate the lateral and longitudinal errors. Shown on the following page is a changing Doppler radar signal developed when the missile changes position in flight.

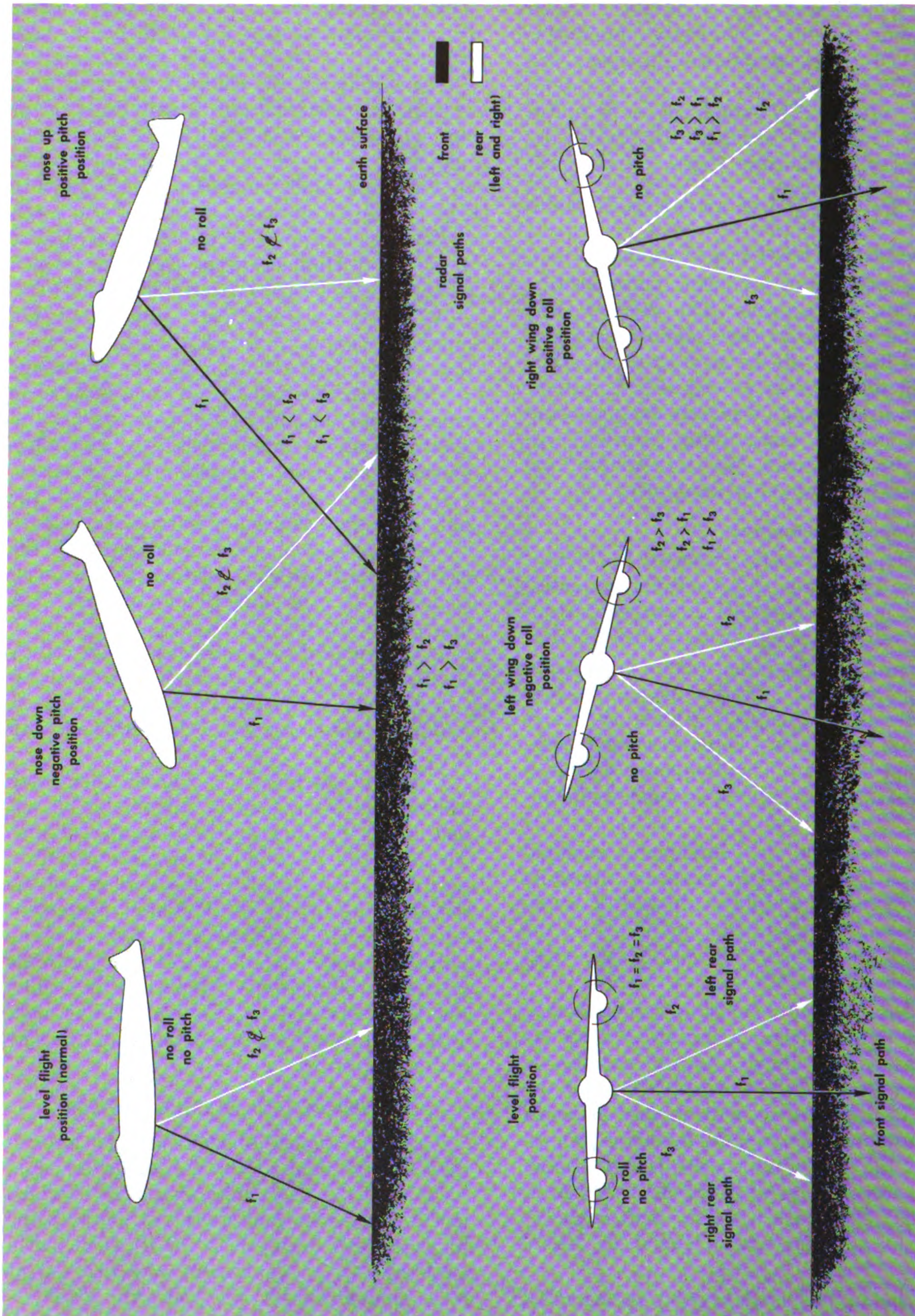
A nose-down (negative) pitching of the missile causes a shortening of the forward radar-signal path and a lengthening of the rear signal paths. This is detected as an increase in true ground velocity if not counteracted instantaneously. A nose-up (positive) pitching is detected as a decrease in true ground velocity.

A left-wing-down (negative) roll of the missile causes a shortening of the left rear path of the radar signal and a lengthening of the right rear path, which is detected as an increase in left drift rate, unless counteracted. The opposite is true of right-wing-down (positive) roll. The vertical gyro compensates for roll by supplying signals to a pitch and roll network, which feeds an elevation servoamplifier to move the antennas about the pitch and roll axis.

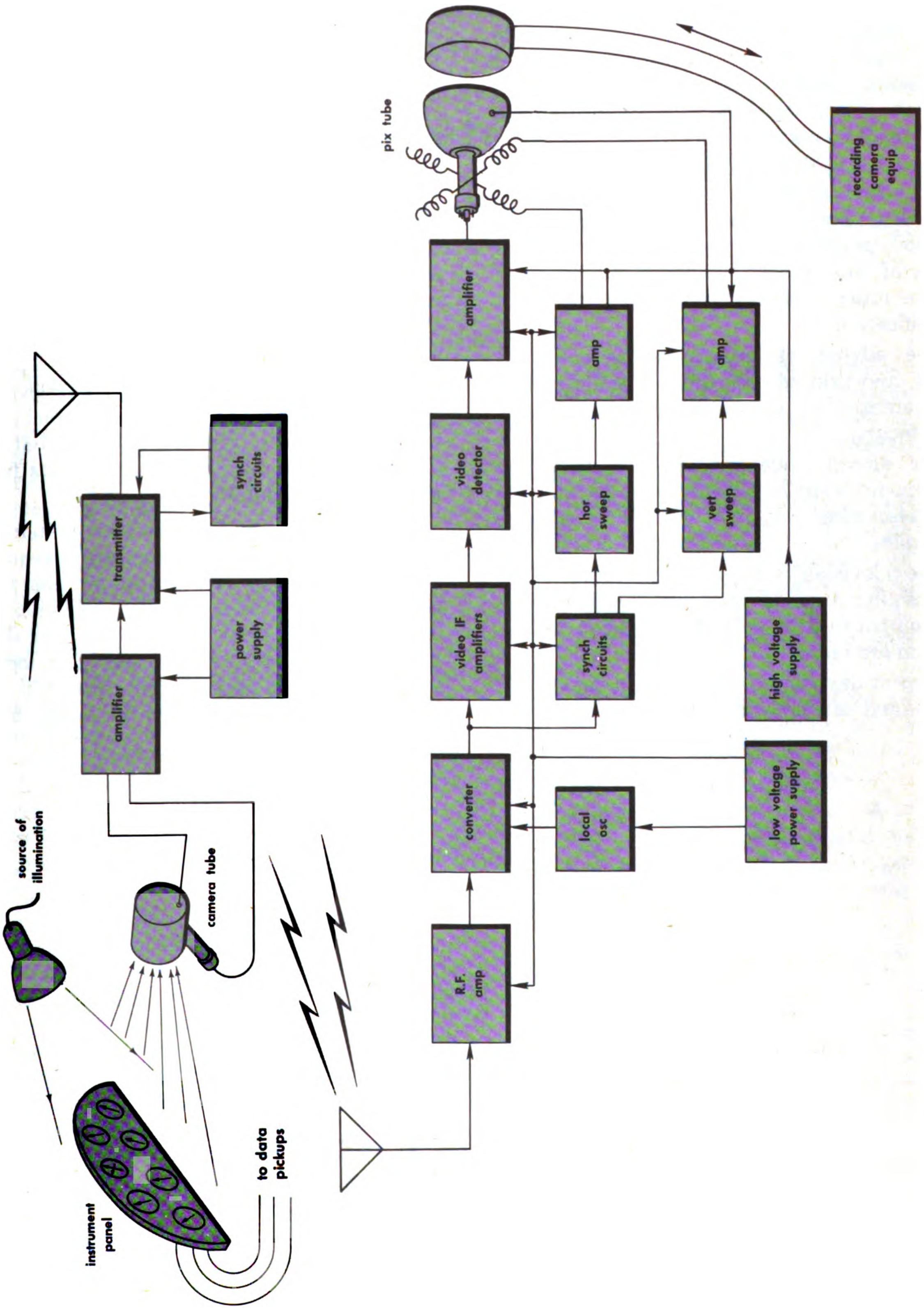
If the missile has a drift angle when pitch occurs, the pitch axis of the missile is not parallel to the pitch axis of the antenna unit. To cause the antenna unit to move about an apparent axis parallel to the pitch axis of the missile requires a combination movement about both pitch and roll axes of the antenna unit. The proper distribution of voltages to the pitch and roll axes is accomplished by feeding the drift voltages into a resolver. The resolver outputs are voltages proportional to the sine and cosine of drift angle.

TELEVISION TELEMETERING SYSTEMS

Television applied to telemetering in aircraft research and flight control is a further development of the earliest methods which involved instrument panels and aircraft structure being photographed during flight. With such methods, the camera equipment was released at some predetermined time from nonpiloted aircraft. This frequently resulted in data being



Doppler radar-signal paths in reference to aircraft position



Television telemetering system (video)

lost when equipment was damaged or lost after release, so some system for televising data was sought.

For large missiles the television system is quite effective. But at the present time, no satisfactory pickup and transmitting equipment has been developed for televising data from small missiles.

However, with long-range telescopic camera tubes, television of missile flight over considerable distances has been accomplished. This method provides for a continuous external check of control surfaces and flight behavior of the missile. It also provides a means of identification.

The advent of transistors, miniaturized tubes, and printed circuits may make missile telemetering by television practical in the near future.

For aircraft telemetering, the television equipment is similar to that used in commercial telecasting, and it operates on the same principles.

The television telemetering equipment does not require a sound channel; therefore, only the components for video transmission and reception are required.

Present-day video systems use an amplitude-modulated carrier, which is susceptible to interference of an AM nature. If the system is used to televise meters and panel instruments, distortion of the picture is not a critical factor as long as the meter readings are visible. However, if structural surfaces are being televised, distortion might be interpreted erroneously as variations of the structure itself. For example, a "jittery" picture might be interpreted as a view of vibrating surfaces, and nonlinearity of the picture might be interpreted as the effect of strain on the surface being viewed.

These latter defects can be overcome by use of positive reading gauges and indicators whose absolute readings are shown on the viewing screen rather than a view of the structure or surface itself.

All indicating instruments in a television system can be mounted on a single panel upon which the camera tube is so focused that the televised view contains continuous data from many channels simultaneously.

When equipment for television systems is developed which meets the necessary requirements for space, power, and ruggedness, telemetering by television will likely come nearest to being the ideal system for collecting data from research aircraft and missiles. But, today, FM/FM telemetering systems are still the most commonly used type in the missile field.

Getting Telemetered Data into Usable Form

This section is concerned with data reduction or, in other words, the methods used for getting telemetered data into usable form. To cover this subject, we'll discuss ground stations, types of antennas, and receiver calibration equipment used in telemetering.

TELEMETERING GROUND STATIONS

A block diagram of an FM/FM telemetering system, greatly simplified, is shown on the following page. The diagram includes both transmitting and receiving sections of the system. In FM/FM telemetering systems, the frequencies of a number of audio-frequency subcarrier oscillators are varied in accordance with the variations of the quantities being measured. These frequencies or tones are mixed and then used to frequency modulate a high-frequency transmitter. This arrangement is represented by the upper half of the diagram.

The signal thus produced is received at the ground station, illustrated in the lower part of the diagram, by a superheterodyne FM receiver which reproduces the mixed subcarrier audio frequencies and then separates them by means of the bandpass filters (discriminators). The individual subcarrier frequencies are converted into DC currents that vary in accordance with the variations of the original quantities being sampled by the pickoffs and transducers in the research missiles.

These currents, in turn, operate galvanometers in a recording oscillograph or other devices which produce a permanent record. The mixed output of the receiver is also recorded on a tape recorder for future study or for insurance against failure of the recording oscillograph. Nominal radio frequencies used are 215-230 mc.

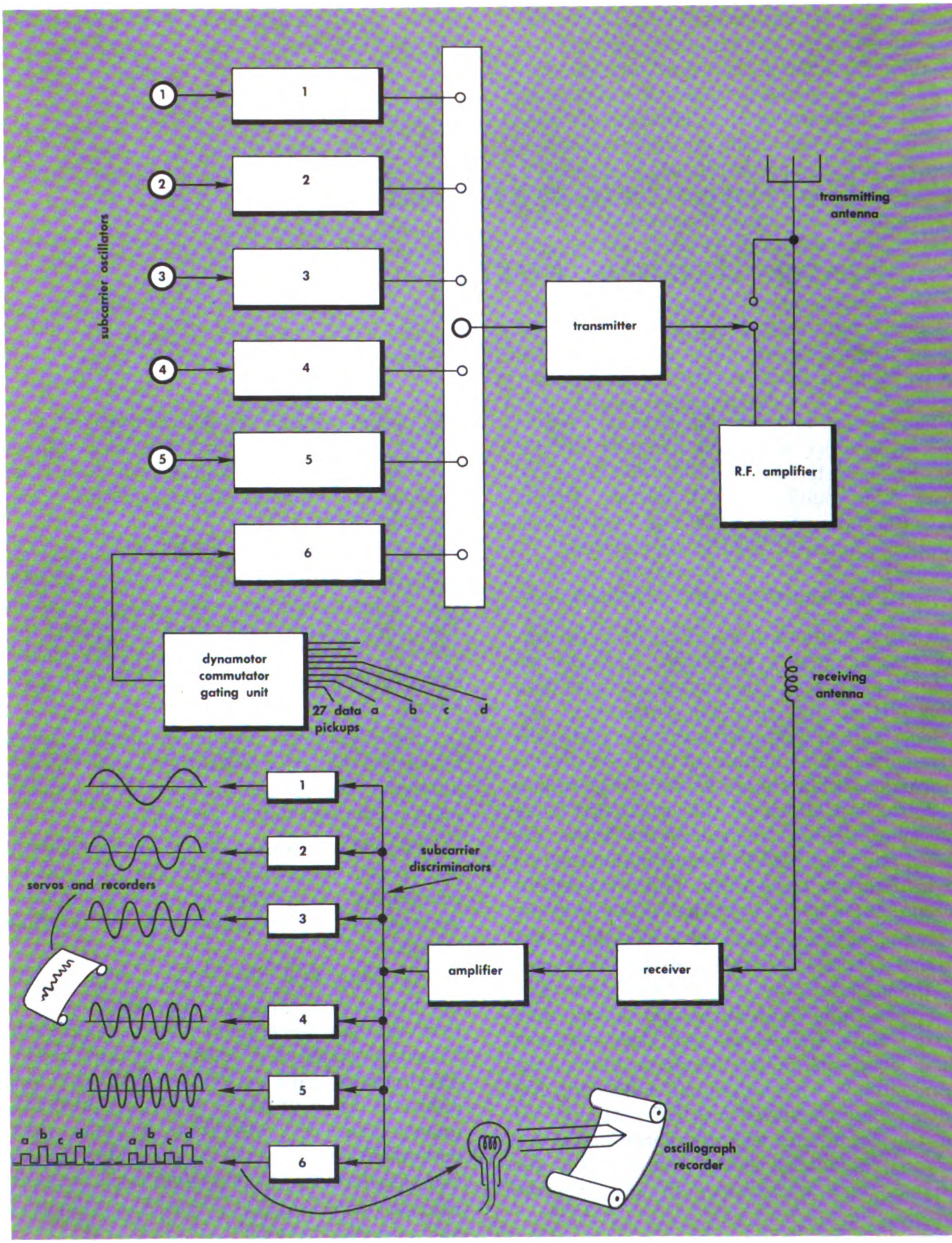
The audio subcarrier-oscillator frequencies employed are in the range of 400 to 32,250 cps. As many as fifteen subcarriers are used. The bandpass of each band is limited to $\pm 7\frac{1}{2}\%$ of the center frequency to insure stability of operation of the subcarrier oscillators.

In some instances, two telemetering systems are employed, each operating on a slightly different RF frequency. Calibration equipment is common to both systems.

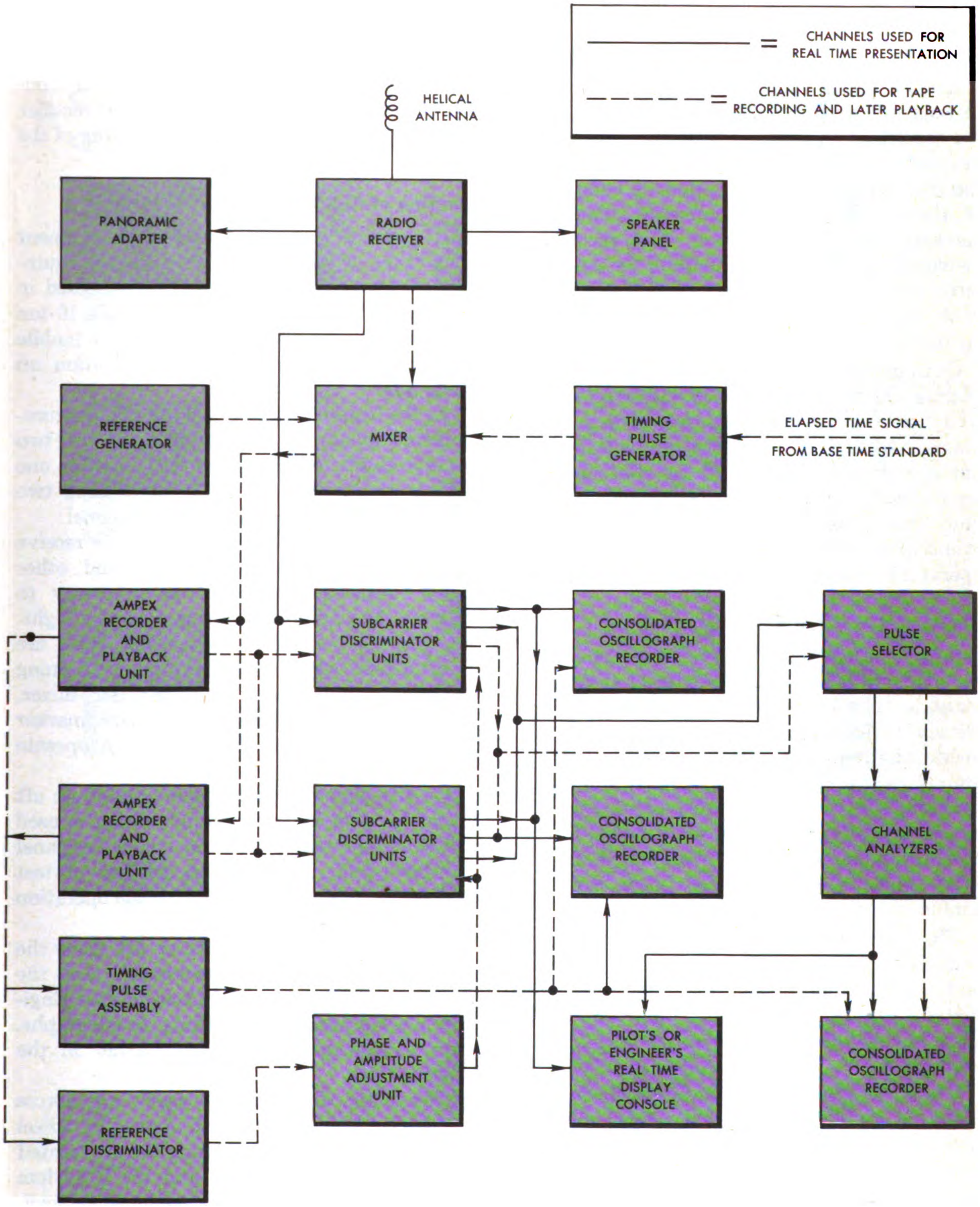
The next illustration shows in more detail, and from a functional standpoint, the major components of a typical ground installation for receiving and recording the telemetering signals.

The signal from the transmitter is picked up by an antenna and conducted to the receiver; there it is amplified and demodulated. Then it is passed on to the system's subcarrier-discriminator units as mixed audio tones. An alternate signal of mixed tones of known frequency from the calibration oscillator (test equipment not shown in the illustration) can be switched into the system in place of the receiver output to calibrate the data-reduction and recording system.

The mixed tones from either the receiver or the calibration oscillator are connected



Typical 32-channel manual telemetering data-separation system



Functional diagram of telemetry receiving and recording station

through the mixer to a tape recorder for *raw-data* storage and to a bank of audio-frequency discriminators, one for each sub-carrier band used.

Synchronizing timing signals are inserted through the timing-pulse generator, providing the recording equipment with accurate time-interval reference. The discriminators separate the channels and convert frequency variations of the bands into DC currents. The DC currents, in turn, operate the recording instruments. The raw data stored on the tape recorder can, at any future time, be played back through the discriminators to produce an oscillograph record.

Outputs of the discriminator units consist of data which is continuous in nature and data which is commutated. One output is a combination of both types. The combination output is desirable for aiding in the selection and separation of a desired tone signal. Note that both continuous and commutated signals eventually reach and are displayed on the operator's console.

Superheterodyne FM Receivers

Superheterodyne FM receivers, such as the modified Navy RBF-3, are used at ground stations. The input impedance of these receivers is set at 52 ohms for the specific operating frequency. The incoming signal is amplified in a two-stage 6AK5 RF amplifier, converted to 10.2 megacycles/sec in a 6AK5 mixer tube by combination with the 6C4 local oscillator voltage, and then amplified in the three-stage 6SG7 IF amplifier. Local oscillator tuning only is employed.

The 10.2-megacycle/sec signal is amplified and clipped in the two-stage 6AC7 limiter, and the constant-amplitude, frequency-modulated square wave is demodulated in a Foster-Seeley discriminator which uses a 6H6 dual diode.

The grid current of the first limiter is metered to determine the strength of the received signal. The DC voltage output of the discriminator indicates correct tuning of the receiver.

The demodulated signal from the discriminator is fed to two cathode followers; the output of one cathode follower may be connected to recording instruments and the other to headphones for aural monitoring. Each cathode follower has an independent gain control. A third cathode follower is connected to the rack output and operates a 500-microampere meter through a meter rectifier bridge to permit continuous monitoring of the receiver level.

Housing of Ground-Station Equipment

At ground stations located on permanent long-range missile proving grounds, the equipment is installed in blockhouses or housed in mobile units, such as the Air Force's 10-ton semitrailer. The floor plan of such a mobile installation is shown in the illustration on page 538.

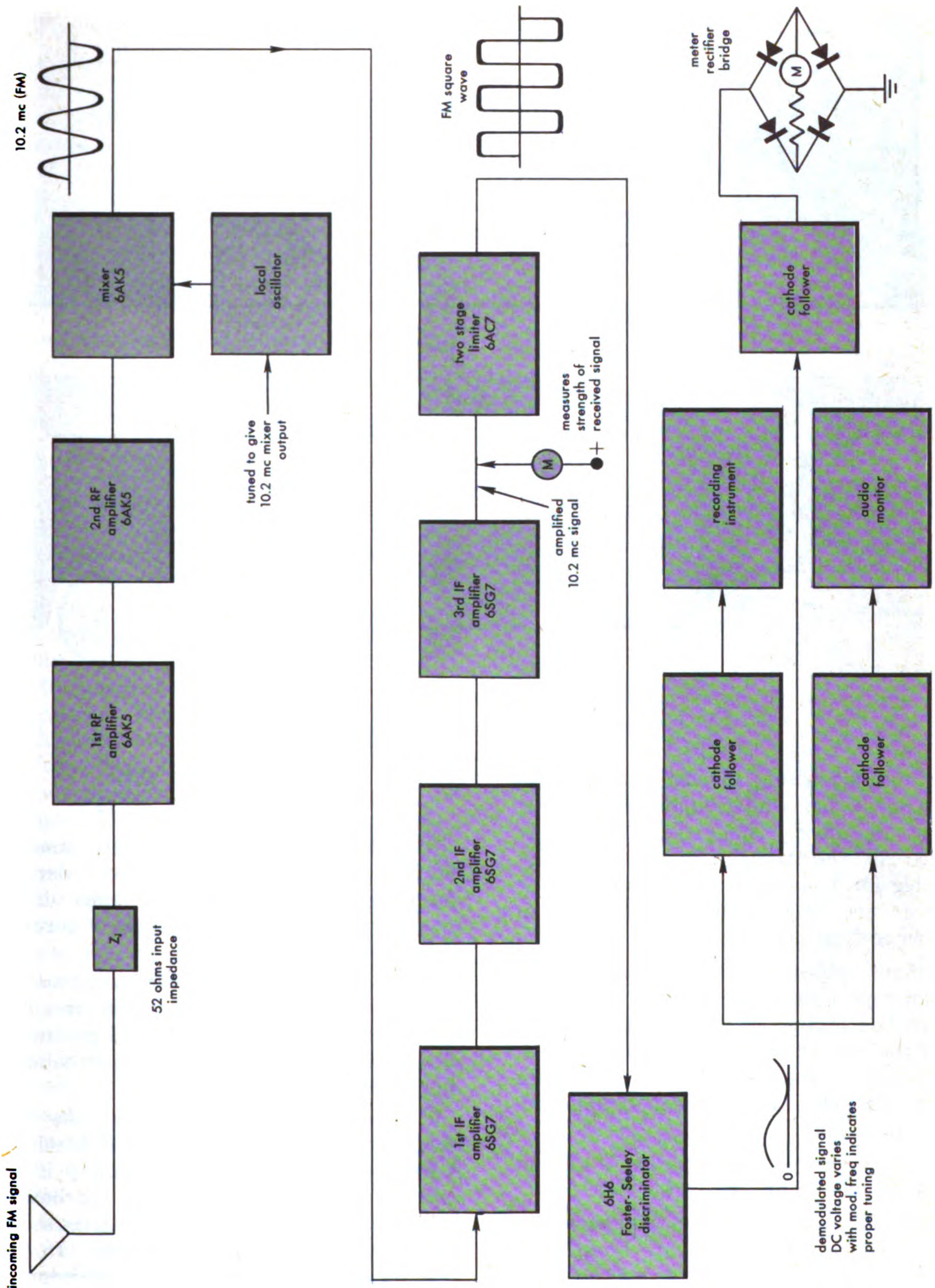
At the front of the trailer is the temperature-control apparatus. Adjoining this are two storage chests, one on either side. Next on one side is the operations table, containing two receivers (A and B) and the control panel.

On this panel are the calibrate-receive switch, the drive-motor switch, and other lights, switches, and controls necessary to operate the equipment during a missile flight. Just to the rear of the operations table are two inclosed standard relay racks containing calibration equipment, tape-recorder mixer, pulse amplifier, two audio amplifiers, marker panel, and a 28-volt DC supply to operate the control circuits.

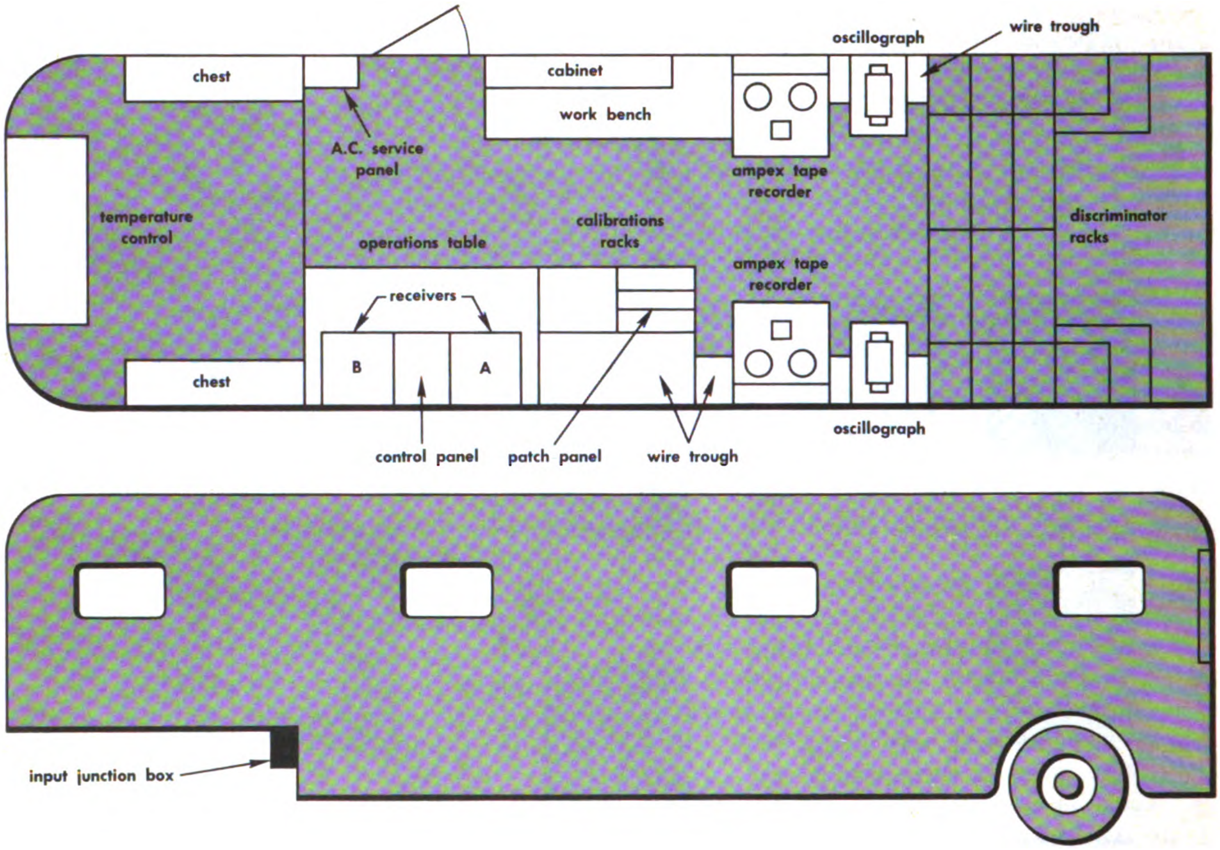
On the opposite side of the trailer just aft from the door is a workbench which is used for field maintenance of the trailer. A panel at this bench terminates a number of test points that are useful in checking the operation of a trailer.

On either side of the trailer aft from the calibration racks and workbench are the Ampex tape recorders and Consolidated Engineering Corporation recording oscillographs. This equipment is described further in the next section.

The main discriminator racks are across the back of the trailer just above the wheel wells. The A system discriminators are located above the left wheel well, and the B system discriminators are above the right wheel well. Their associated power supplies are in adjacent racks, together with Ampex Capstan motor amplifiers.



FM telemetering receiver



Floor plan of telemetering trailer

All interunit wiring in the trailer is run through 3″ x 10″ wire troughs along either edge of the floor. The cables are loosely installed in the troughs and are readily accessible for maintenance by removing the trough covers. Termination of all wires is in a telephone-type junction box.

All input cables including telephone lines and antenna inputs connect to the input junction box located on the outside vertical face of the *step* of the trailer. The box is on the left side of the trailer and conveniently connects through to the wire trough under the operations table.

The 220-V, 60-cps AC inlet plug is located similarly on the right side of the *step*.

All telephone lines, together with pertinent internal inputs and outputs, are brought to the patch panel located on top of the rearmost calibration rack. Special, as well as normal, interconnections of equipment are possible at this panel.

ANTENNAS USED WITH GROUND RECEIVER UNITS

Various types of antennas are used in different ground-station installations, depending somewhat on the type of antenna utilized by the transmitter and the angle of coverage of the radiated signal.

However, since *helical-beam antennas* are quite efficient for FM/FM carrier reception and are widely used with mobile ground receiver units, the discussion here is centered on this type.

This antenna was designed at the New Mexico College of Agriculture and Mechanic Arts. It consists of a helix operating in the *axial mode* of radiation; that is, the helix is directed so that its axis is in line with the source of radiation (transmitter antenna). In this mode of operation, the field is a maximum in the direction of the helix axis and is circularly polarized, or nearly so. The decibel gain of the antennas is about 11 when receiving from a

circularly polarized antenna and about eight when receiving from a linearly polarized antenna. The beam width between half-power points is about 36 degrees.

A field-intensity meter is incorporated in the antenna circuit so that the antenna can be oriented for maximum signal strength.

A helical antenna has a characteristic impedance of approximately 140 ohms and is matched to a 52-ohm, RG-8/μ coaxial cable by impedance matching sections. The connection on the antennas, cables, and matching sections are type-N constant-impedance fittings.

When a trailer is used to house the receiving equipment, the antenna cables are attached to coaxial fittings in the input junction box of the trailer.

RECEIVER CALIBRATION EQUIPMENT

Calibration of a telemetering receiving system is accomplished by substituting for a received signal of mixed audio frequencies a number of signals of known frequency from the calibration equipment.

An oscillator for each band is designed to produce any one of three frequencies. The desired frequency can be selected by a switching arrangement. The three frequencies used are the channel-center frequency and the upper- and lower-limit frequencies for the channel (center frequency $\pm 7\frac{1}{2}\%$).

When the frequencies are recorded in time sequence, a three-point calibration is obtained. The telemetering equipment produces data that is linear enough so that no other calibration is required. By means of ganging the switches from all calibration oscillators, simultaneous calibration of all channels is obtained.

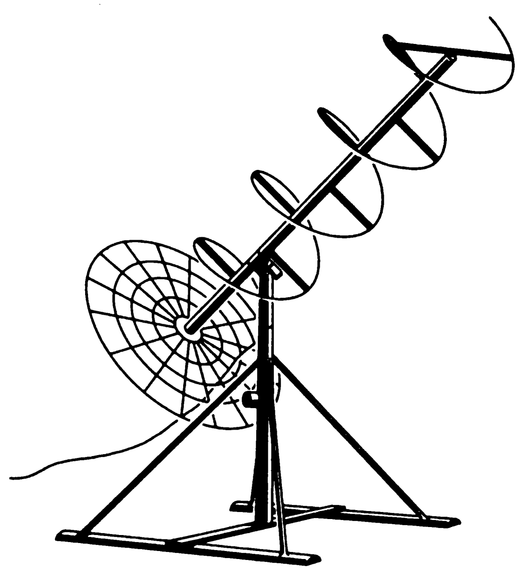
The output (0.4 rms) of each calibration oscillator is fed into a combination mixer and amplifier stage, as shown on next page. Mixed tones of center frequencies, or the band limit frequencies of all channels, are fed to the discriminator racks when the calibration switch is in any one of the three calibration positions.

A bandpass filter in each discriminator separates its particular channel frequency from the mixed input. This signal is applied to an instrument which measures events per unit time by counting the cycles generated by the oscillator during a 1- or 10-second period of time, depending on the required degree of accuracy. This permits the frequency of each oscillator to be accurately adjusted.

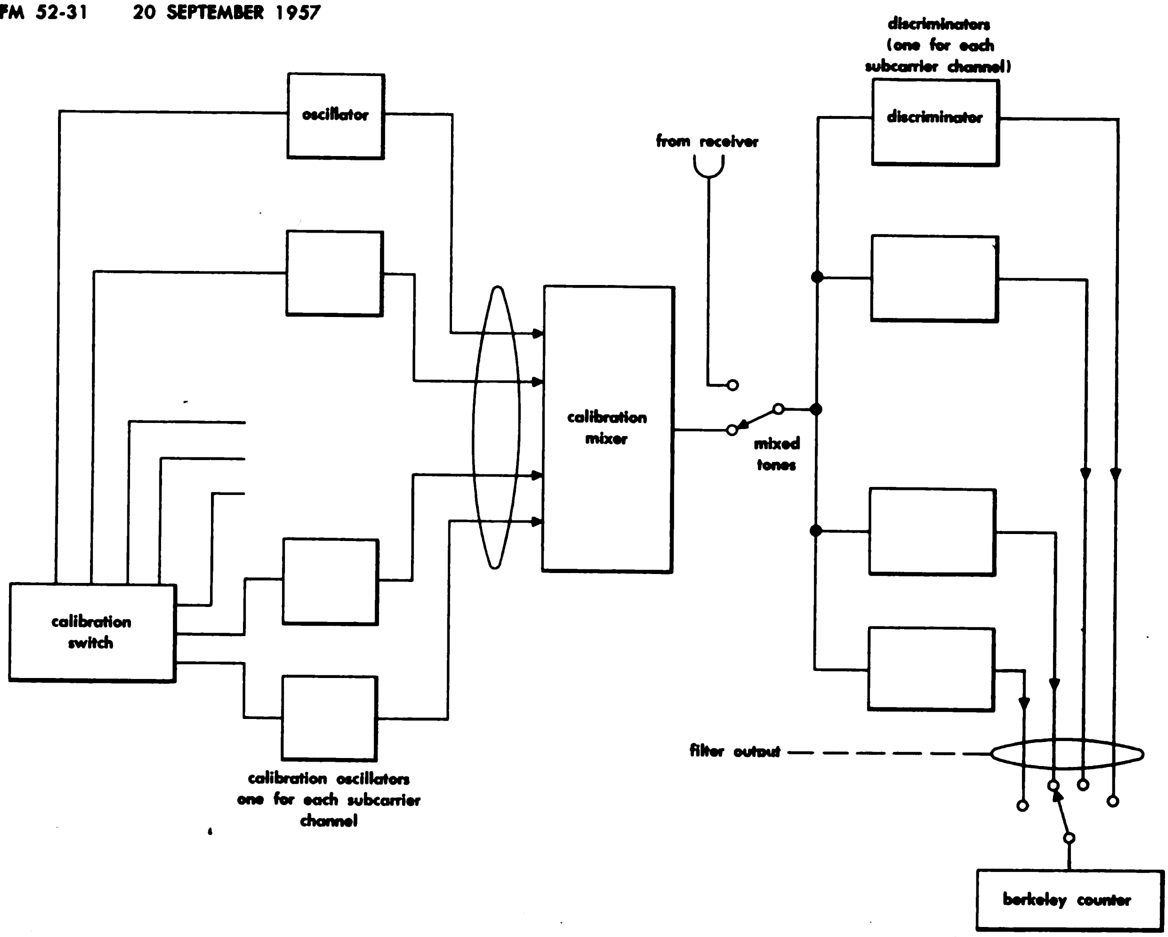
The Berkeley Universal Counter and Timer, used in the calibration procedure, is a cycle counter. Each calibration frequency is measured by counting cycles during a 1- or 10-second interval. The accuracy of the instrument is $\pm$ 1 count. The counting interval is chosen so that the ± 1 count will always be an error of less than $\frac{1}{2}\%$ of a $\pm 7\frac{1}{2}\%$ subcarrier band. Time interval and frequency ratio can also be measured by the Berkeley counter.

Calibration Oscillators

In the telemetering calibration system of the Jet Propulsion Laboratory, the calibration oscillator chassis provides space for fourteen plug-in oscillators, as well as for circuitry for mixing the outputs of all fourteen oscillators. This chassis is schematically drawn on page 541. Here, the mixed tones are amplified and passed through a cathode follower output tube. The output level for all fourteen mixed tones is set at 1.5 volts rms. This output voltage is measured by a built-in, vacuum-tube voltmeter. The power supply delivers 200 volts at 200 milliamperes regulated power to the plug-in oscillators and the rest of the circuit.



Helical antenna used with telemetering receivers



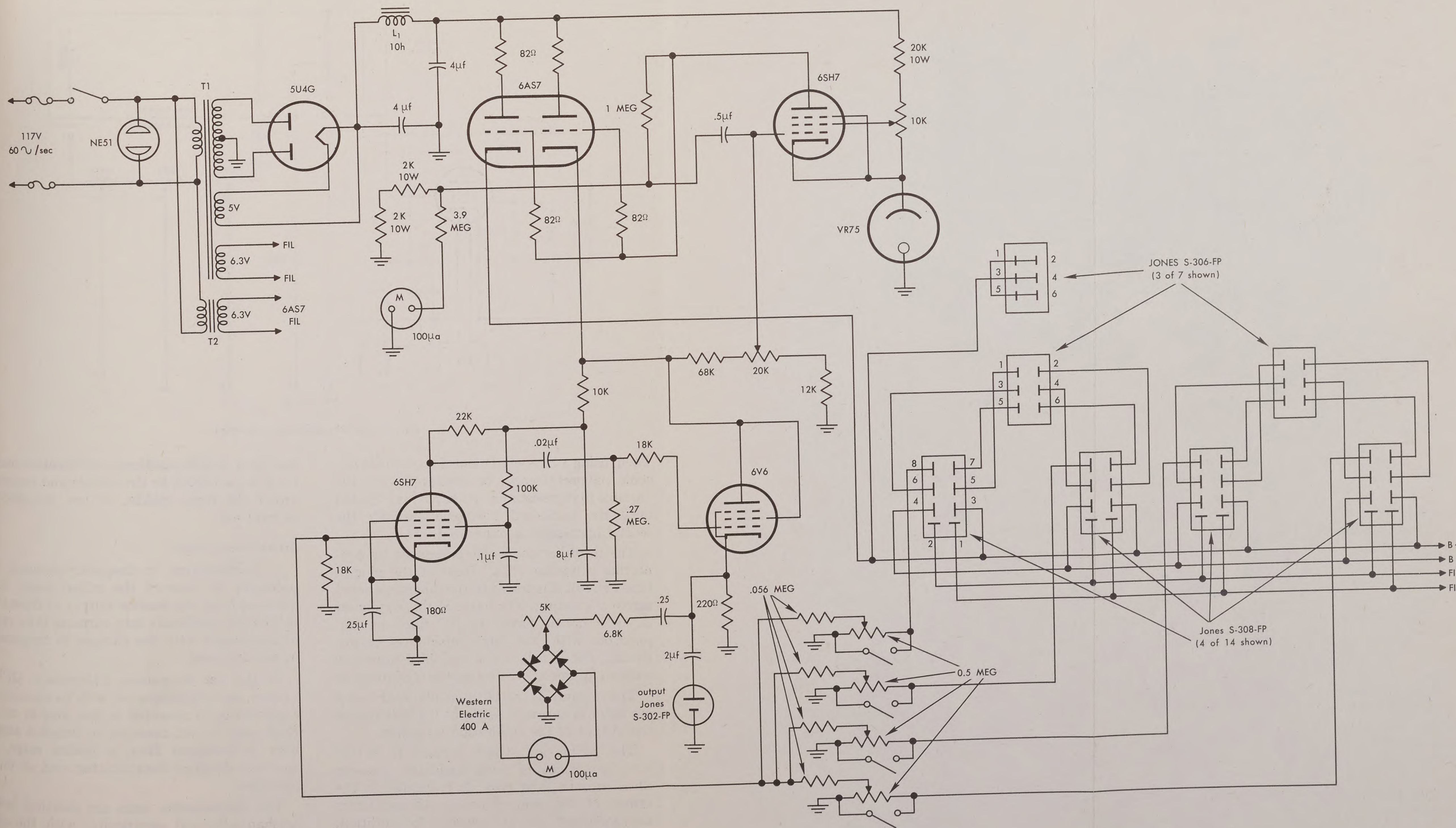
Telemetering receiver calibration equipment

The calibration oscillator for each channel is built on its own plug-in chassis. Electrically and mechanically, the oscillators are identical except for a tuned circuit. Each tuned circuit is a plug-in unit of its respective oscillator. This design provides for a quick change of units in case of a failure. A spare unit for each channel is always kept available, thus making it unnecessary to disable the entire ground telemetering equipment in order to repair a single oscillator.

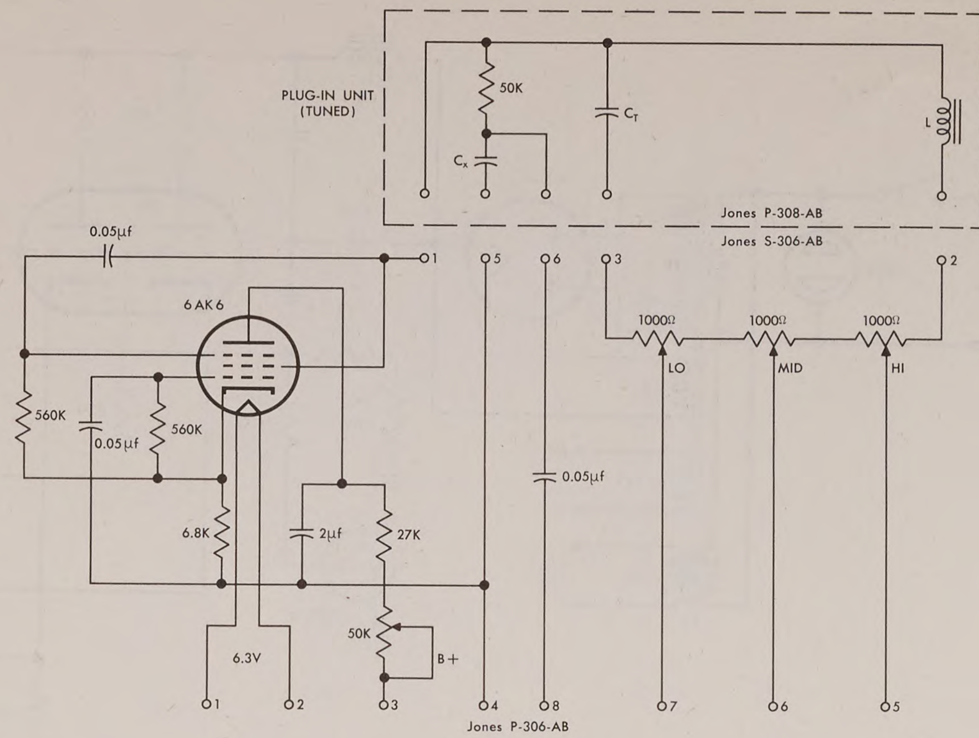
A modified transitron circuit is used in each oscillator. This tuned circuit has proven to be quite stable when supplied with regulated filament voltage and 200 volts of plate supply. When properly adjusted, it is relatively free from distortion. This tuned circuit is an RLC resonant type having a resonant frequency that can be varied over wide limits depending on the amount of resistance employed in each branch. Frequency variations of $\pm 7\frac{1}{2}\%$ are easily attained. The three 1000-ohm potentiometers provide an easy way to set the frequency to the HIGH, MIDDLE, and LOW values as required.

As in all oscillators, harmonic distortion appears in the output wave form because of the amplitude-limiting action of the tube. Second harmonics (and, to a lesser degree, third harmonics) are undesirable in an FM system because such harmonics produce undesirable beats in higher-frequency channels. Both even and odd harmonics exist in the transitron oscillator. Either one can be greatly reduced by adjusting the value of the plate load resistor so that the oscillator operates around the proper portion of its *negative resistance* characteristic.

This adjustment is performed at "midband" setting (center-frequency), each oscillator being checked completely independently of all other oscillators; to do this, all the oscillators except the one being checked are disconnected. The plate supply is set to exactly 200 volts.



Calibration oscillator and amplifier chassis



Plug-in unit of calibration oscillator

Then, using a Hewlett-Packard wave analyzer, sonic panoramascope, or similar device, the variable plate resistance (50-K pot) is adjusted until the undesired harmonic (usually the second harmonic) is minimized.

The third harmonic also is reduced by connecting condenser "Cx" (part of the plug-in tuned circuit illustrated in the diagram above) across the output. The value of this condenser is such that it forms an RC filter in conjunction with the high-impedance output circuit, and it attenuates the third harmonic without appreciably affecting the fundamental.

The output level of a single calibration oscillator is normally set to be 0.4 volt rms at the output of the calibration amplifier.

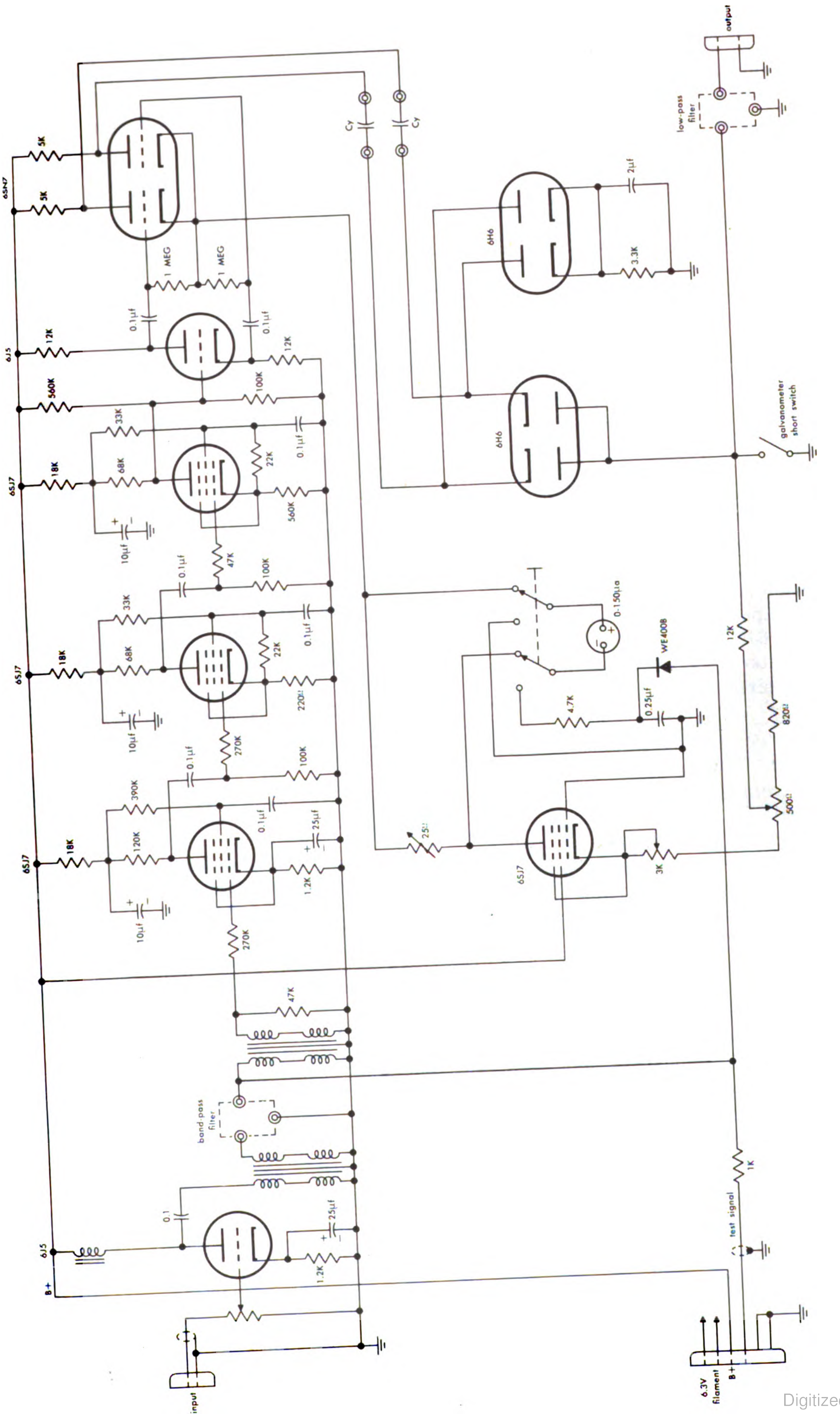
The calibration switch is used to switch the frequencies of each oscillator between HIGH, MIDDLE, and LOW. It is located in the center of the control panel. All oscillators are switched simultaneously. In addition, this switch, which is a twenty-pole, five-position rotary switch, connects the receiver to the system. The receivers are connected when the switch is in either end position. In the three middle positions, a calibration oscillator is connected to the system and supplies either the high, middle, or low frequencies as required.

Discriminator Units

A discriminator, or frequency-detector, is necessary to convert the mixed audio frequencies from the receiver output or from the calibration oscillators into currents that vary in accordance with the changes in frequency in each channel.

In the Jet Propulsion Laboratory (JPL) system, each discriminator, with its associated components, is mounted in one plug-in unit. Each plug-in unit contains an isolation amplifier, a bandpass filter, a limiter stage, a specially designed discriminator, and an output filter.

The discriminator units are identical both mechanically and electrically, with the exception of the frequency-sensitive components (bandpass filter, two discriminating condensers, and low-pass output filter) which are separate plug-in units. As in the case of oscil-



Low-frequency discriminator unit

lators, this design permits rapid replacement of defective units without shutting down the entire system. Also, a discriminator plug-in unit can be readily adapted to any channel by inserting the proper frequency-sensitive components.

Performance of these units has been satisfactory. Their total frequency drift from all causes is less than 0.5%, and their nonlinear characteristics are less than 0.25%.

With lead impedance between 0 and 1500 ohms, the unit operates like a constant-current generator with an internal impedance of 15,000 ohms. This arrangement insures good linearity over a wide range of impedance variation in the input and output circuits. The arrangement also permits the use of many different types of recording instruments without special impedance-matching units and with negligible variation between the unit's indicated outputs.

As shown in the diagram on page 543, the circuit of this low-frequency discriminator unit comprises an input amplifier stage using a 6J5 triode. The 6J5 triode isolates the following bandpass filter from the input line that supplies the mixed subcarrier audio frequencies.

A potentiometer in the grid circuit of the 6J5 sets the level of the input signal.

The output from the 6J5 is fed through a transformer that matches the plate-resistance of the tube to the 500-ohm input impedance of the bandpass filter. Since this transformer is not capable of carrying the plate current of the tube, the stage is shunt-fed through a choke. The choke is used rather than a resistor in order to obtain a greater amount of gain (voltage-amplification) from the stage.

The bandpass filter is a specially built *lumped-constant* filter, which operates at a 0-db level (6 milliwatts in a 500-ohm load) with an insertion loss of less than 3 db at minimum attenuation. The response is flat within 3 db over the bandpass with at least 30-db attenuation at the near edge of the adjoining bands. This means that, within the band pass, the attenuation of any channel frequency between $\pm 7\frac{1}{2}\%$ of the mid-frequency is never greater than one-half. However, frequencies just outside the band limits may be attenuated to one-thousandth of their normal level.

To show that 30 db represents a ratio of 1000-to-1 in power:

$$30 \text{ db} = 10 \log \frac{P_1}{P_2}$$

$$\log \frac{P_1}{P_2} = 3$$

$$\frac{P_1}{P_2} = 1000$$

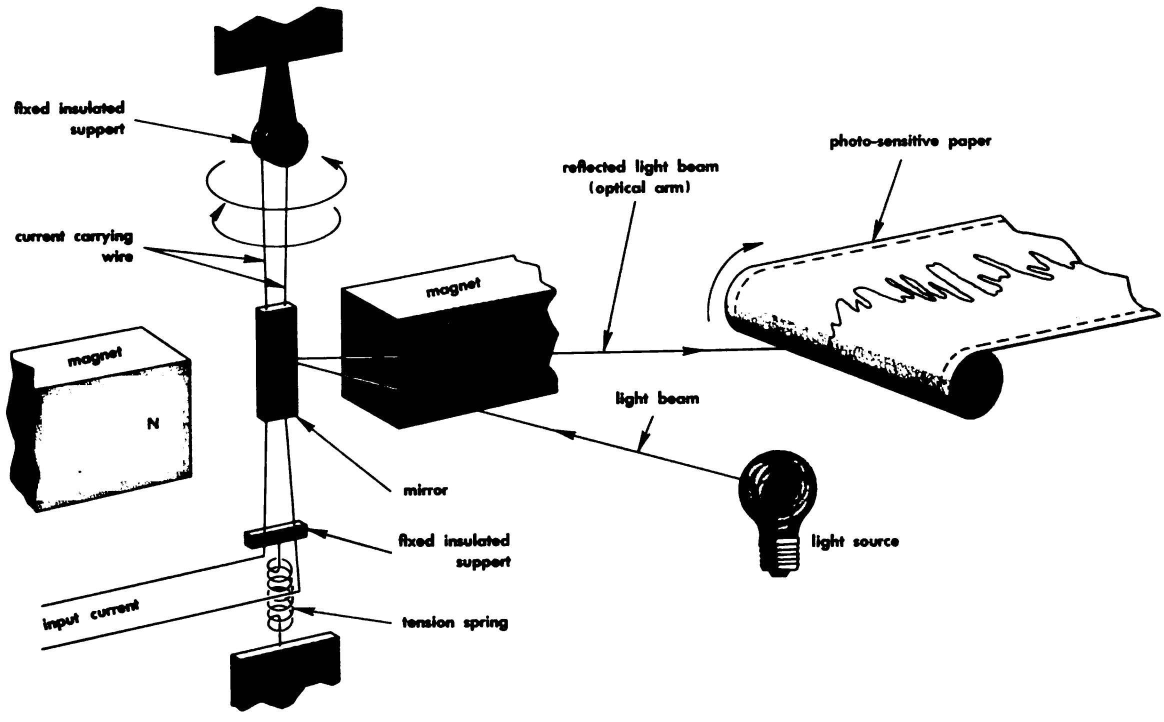
Thus, 30 db means that P_1 is 1000 times as great as P_2, or P_2 is $\frac{1}{1000}$ of P_1. A 30-db attenuation means an output power of $\frac{1}{1000}$ of the input power.

The output of the filter is matched to the grid of the following stage (6SJ7 limiter) through a 500-ohm line to the grid-matching transformer.

The selected audio-frequency signal which normally is set at a level of 0.4 volt is amplified by three limiter stages employing 6SJ7 sharp cut-off pentodes operating as overdriven amplifiers. The overdriven amplifiers amplify and clip the sine-wave output of the filter into square waves. The gain of these three limiter-amplifier stages is sufficient to produce a square wave output from any input signal of from about 0.02 volt to 2 volts in amplitude.

Following the limiter stages is a new circuit based on a constant current system. Basically, it is a current-generating circuit that is sensitive to variations in frequency and possesses certain advantages over the ordinary Foster-Seeley discriminator. The circuit employs a 6SN7 dual triode which operates as a push-pull electronic switch. Output of the limiters is inverted at the plate of the 6J5 triode driver and phase-inverter (paraphase inverter) tube and is coupled to the grid of one section of the 6SN7 push-pull switching tube. The same voltage, but 180° out of phase, is coupled from the cathode of the 6J5 to the grid of the other section of the 6SN7.

A 6SJ7, used as a constant-current tube, limits the current through each section of the 6SN7 dual triode to any value determined by the setting of a variable cathode-bias resistor



Bifilar recording galvonometer

(2.5-K or 3-K potentiometer) in the cathode circuit of the 6SJ7.

Each section of the 6SN7 alternately conducts a current of constant amplitude. Condensers (Cy) in the plate circuits of the 6SN7 are alternately charged and discharged by a current of constant peak value. The condenser charging current consists of a series of discrete pulses, each of constant average current. The number of pulses per unit time is a linear function of frequency; that is, an input of a certain frequency produces a specific number of pulses in a unit of time:

$$N = \frac{f}{T}$$

The current pulses are rectified by a bridge rectifier employing two 6H6 dual-diode tubes, and the output is filtered and applied to a galvanometer or other indicating device.

The output of the bridge consists of a large DC component upon which is superimposed the small current variations representing frequency.

In order to operate the galvanometer near mechanical zero, a bucking current is obtained from a 500-ohm resistor in the cathode circuit of the 6SJ7 constant-current supply tube and is connected through a 12,000-ohm resistor to the galvanometer to oppose the bridge output.

This circuit is across the bridge output and therefore imposes a load on the bridge. Thus, it is necessary to keep bucking-circuit resistance high if good current regulation is to be maintained. Since the bucking current is part of the constant-current loop, common variations of the bucking-current and signal-current average are equal and opposite. Thus, *zero* position of the galvanometer is unchanged. In order to operate the galvanometer at ground potential, the bucking current is obtained from the cathode circuit of the 6SJ7 constant-current tube rather than from the plate. Since the screen-grid of this tube draws current which is not a part of the regulated current, there is a slight variation between the signal current taken from the plate and the bucking taken from the cathode, but the difference is negligible because the amount of screen current is small.

The bridge rectifier uses 6H6 tubes rather than crystal diodes to eliminate large current drifts caused by changes in the forward and backward resistance of crystal diodes. Changes in forward and backward resistance of crystal diodes occur when the diodes are operated under varying temperature conditions. Vacuum tubes of the 6H6 type are free from such effects but do have high sensitivity to filament voltage when used in this manner. This sensitivity is caused by the bucking current flowing through the bridge in parallel with the galvanometer during part of the cycle.

As resistance of the bridge changes due to variations in filament voltage, galvanometer current changes, causing drift.

This tendency is overcome by biasing the ground end of the bridge to about 1 volt by means of a 2-μf capacitor and a 3300-ohm resistor. This voltage is sufficient to prevent the bucking current from flowing through the bridge. Similar biasing of a crystal-diode bridge would not affect the temperature drift.

A 150-microampere meter is normally connected in the constant-current circuit. The meter serves as a current monitor, and its reading is an indication of the sensitivity of the channel.

When the switch button under the meter is depressed, the switch is closed, and the meter reads the signal level at the bandpass filter output. This voltage is also brought out through the power plug for the purpose of checking frequency and wave form.

A galvanometer shorting switch is also included in the output of the discriminator unit to facilitate oscillograph adjustments.

Plate supply for the discriminator unit is a rack-mounting Lambda model 28 commercial power supply. This supply furnishes a regulated output of 300 volts at 100 milliamperes. One power-supply unit furnishes voltage for two discriminators.

The discriminator filament voltage is critical. If less than 6 volts is used, the output of the discriminator is affected appreciably by filament-voltage changes. Over 6 volts, the discriminator is quite insensitive to small voltage changes. In order to insure adequate filament voltage, a separate 7.5-volt transformer for each discriminator is used to maintain at least 6 volts at the tube terminals. The voltage is set to 6 volts by means of a rheostat in the primary of each transformer.

To further insure filament-voltage stability, a Sorenson voltage regulator is used at the 110-volt AC source.

Controlling Galvanometer Response

Flexible data presentation requires a means of controlling galvanometer response. The method described here is, in particular, the method of matching the output of the telemetering discriminator through filters to the galvanometers. Consolidated Engineering type 7-212 galvanometers, or similar types, are used.

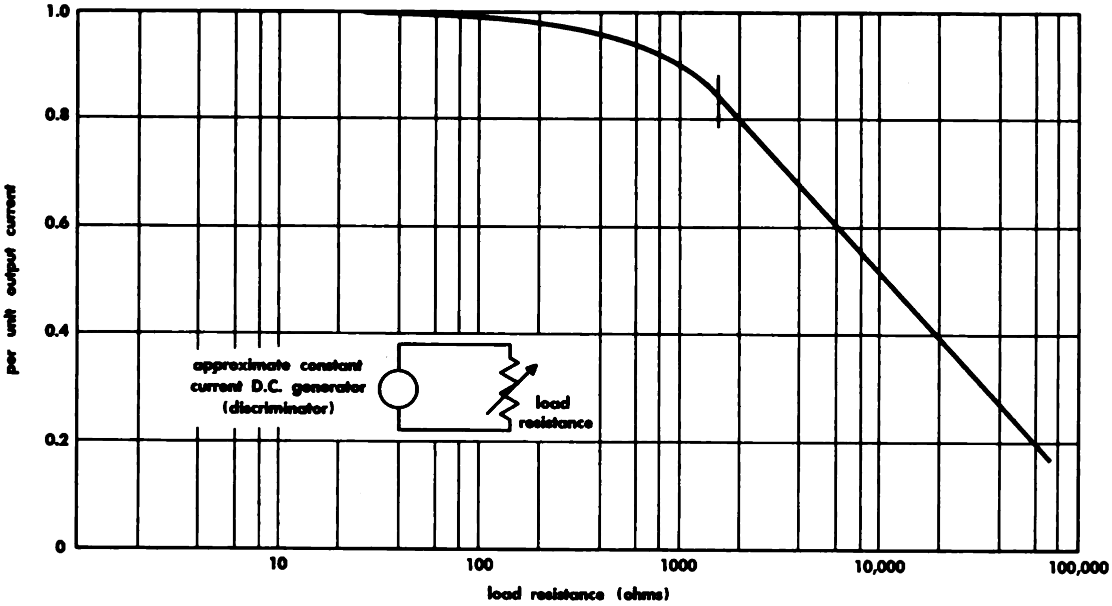
The matching method serves to improve the cut-off characteristics of the galvanometers and control the system frequency response without affecting DC response or amplitude sensitivity of each channel.

Discriminators, oscillographs, and galvanometers of different makes and types possess different characteristics, but the basic matching principles apply in all cases. Only the values of the filter-matching circuit components vary to suit the requirements of the equipment employed.

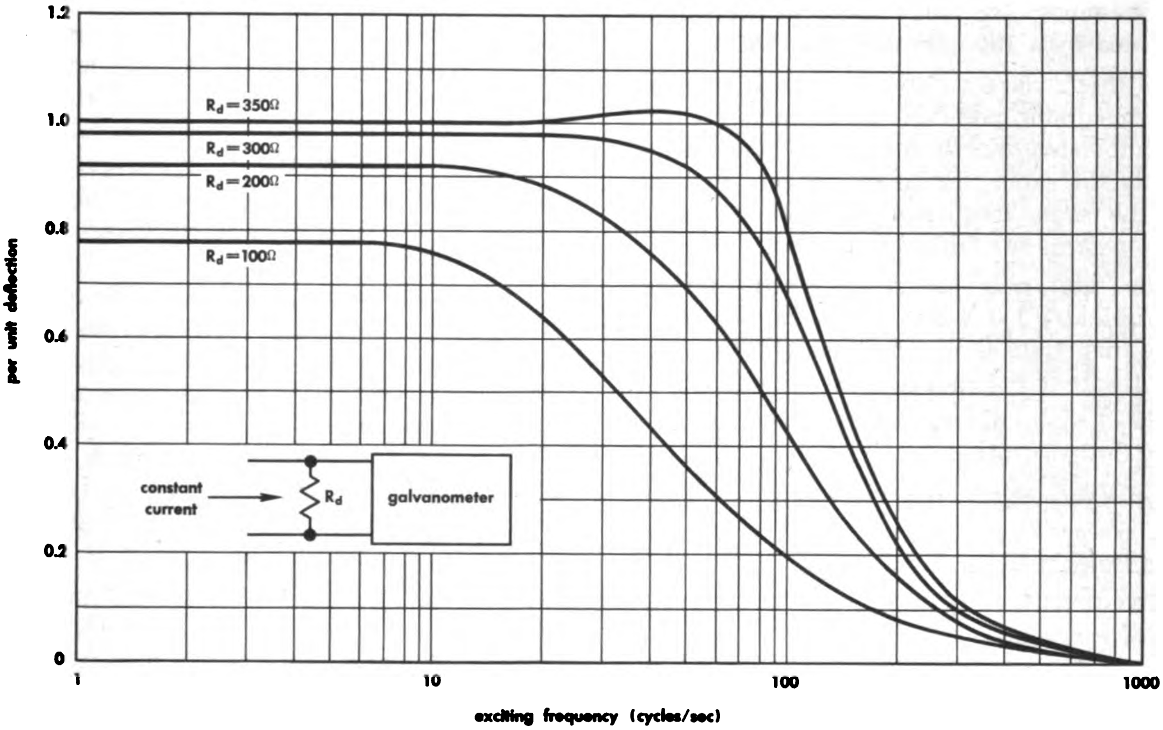
Considering the discriminator as a constant-current (or constant-voltage) generator whose current is modulated by its input frequency, the external characteristics of the discriminator with constant input frequency can be represented as shown in the graph above right.

A change of frequency raises or lowers the entire curve. As shown in the graph, the load can be as high as 1500 ohms without drastically lowering the output current. If greater resistance is used as a load, the output may become nonlinear with input frequency due to the effect of reflected impedance on the source.

Characteristics of a Consolidated Engineering type 7-212 galvanometer as shown in the next graph indicate that maximum flat response is obtained to about 60 cps with a 350-ohm damping resistor. Reduced frequency response can be obtained by lowering the damping resistor to not less than 100 ohms. Greater damping does not cut response much, but greatly reduces its DC sensitivity.



Discriminator output characteristics with constant input frequency



Characteristics of consolidated-type 7-212 galvanometer with damping resistance to a constant current

The curves have characteristics similar to those of a low-pass filter, but they allow too much response at 500 to 1000 cps. Unfortunately, noise in the form of the discriminator input frequency and its second harmonic appears at the output of the discriminator. These frequencies are around 500 and 1000 cps for the lowest telemetering channel.

This necessitates improvement of galvanometer cutoff characteristics, because secondary resonances (harmonics) appear at much higher frequencies then 500 and 1000 cps and these resonances cause some response at certain spots in higher telemetering bands unless removed by other filtering.

A filter connecting the discriminator and the galvanometer has three purposes:

1. To improve the cutoff characteristics of the galvanometer in the lower three bands.
2. To filter out all frequencies that might excite the galvanometer at any of its secondary resonance points.
3. To reduce the response of the galvanometer when it is desired to cut response of the galvanometer below that which can be obtained by a 100-ohm damping resistor.

All these characteristics should be obtained with maximum possible output current to the galvanometer. Furthermore, all filters must supply the same current to the galvanometer for the same frequency deviation within a band regardless of band frequency.

The filter must be so designed that the discriminator looks into an impedance of not more than 1500 ohms, while the galvanometer sees the required damping resistance.

These requirements can be met by a properly designed filter similar to the generalized filter circuit illustrated below.

The design of filters is beyond the scope of this manual. Their specific function in this case is similar to that of an impedance matching device which has desired frequency-cutoff characteristics.

Auxiliary Equipment for Data Reduction

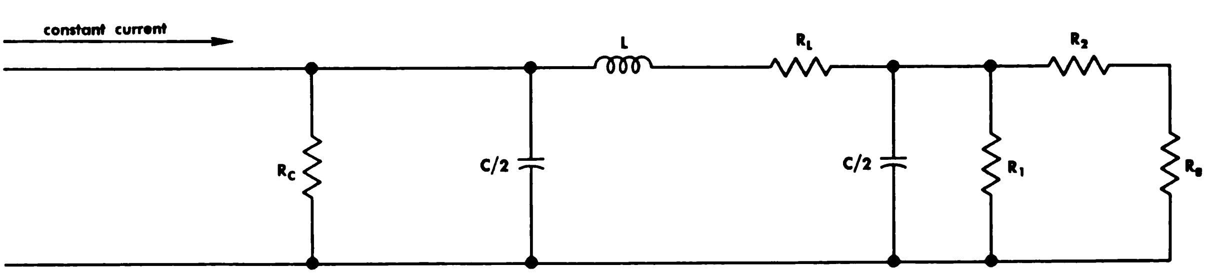
In addition to the galvanometer recording oscillographs and tape recorders, auxiliary equipment and circuits for data reduction and evaluation are employed.

DETERMINING FIELD STRENGTH OF RECEIVED SIGNAL. Knowing the field strength of a received signal is useful in determining if the system is functioning properly. For this purpose, the grid current of the first limiter tube is available. This current is passed through a milliammeter mounted on the antenna pedestal or near the antenna so that the indication of the meter can serve as a guide in positioning the antenna.

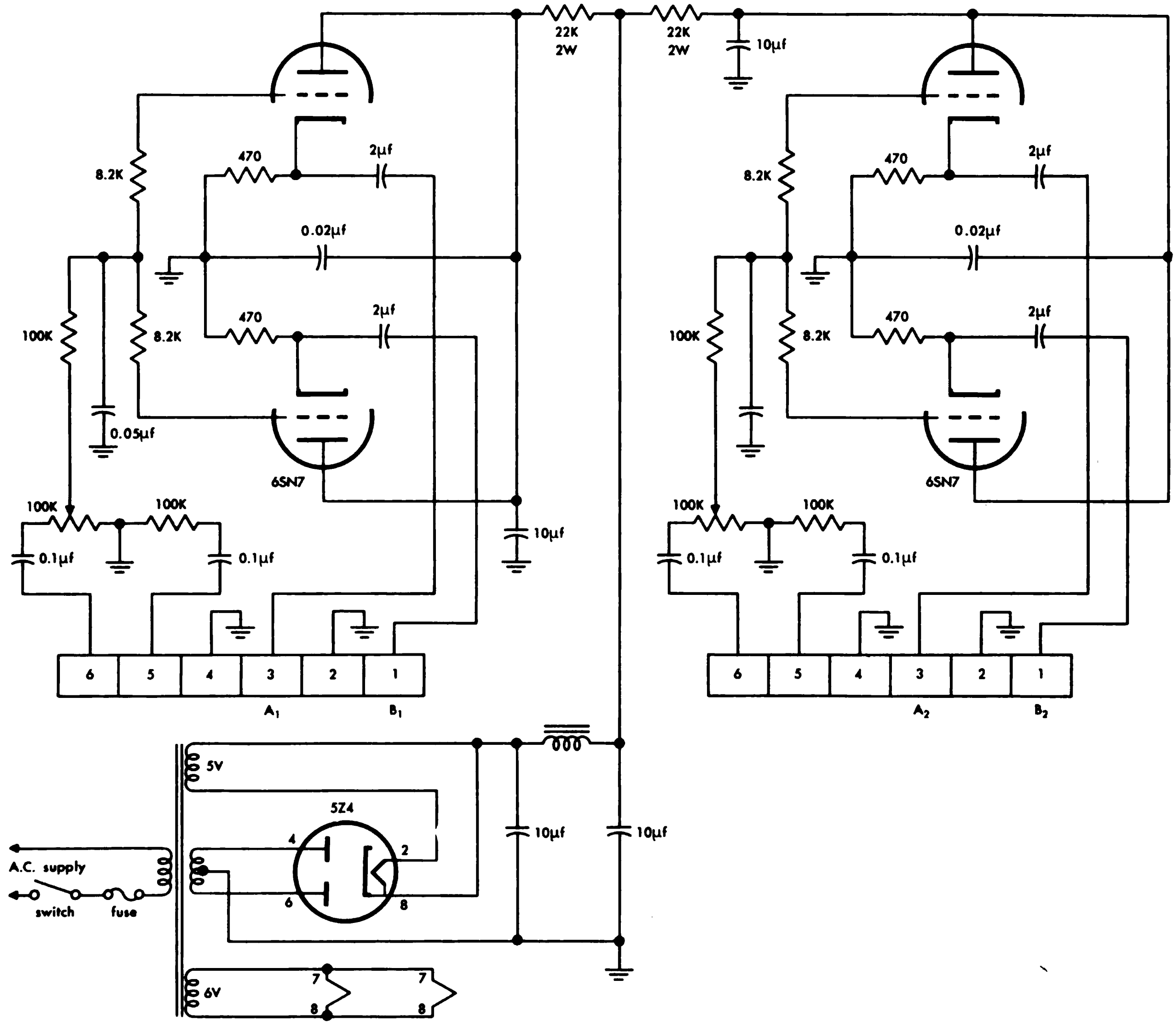
After proper attenuation, the current is passed through a galvanometer in the Consolidated Engineering oscillograph for recording along with other data.

Calibration of the field-strength trace is accomplished by inserting known levels of signals from a suitable signal generator into the antenna plug of the receiver. The galvanometer attenuator is set to give adequate deflection with high signal level. Then, various levels are recorded to provide deflection-vs-signal strength data.

PROVIDING TIMING SIGNALS. The ground stations at a test range may provide timing signals for the use of all the participating agencies. Generally, ½-second pulses and 10-second coded pulses are available. Both of these pulses can be recorded on each oscillograph.



Generalized constant-k filter and galvanometer circuit



Wiring of a pulse amplifier

MAKING TIME MARKS MORE READABLE. In order to make the time marks more readable at the slow paper speed used, the pulses are integrated or lengthened in a pulse amplifier, schematically shown above.

This amplifier also serves the purpose of isolating the galvanometer from the telephone line over which the timing signals are transmitted. Thus, a large voltage across the line caused by lightning or other reasons cannot reach the galvanometer at sufficient level to damage it.

In addition to the external timing, 6-volt low-impedance pulses occuring once per second can be supplied as a record marker for all data including those recorded by other than the telemetering means. These pulses are used to indicate firing and takeoff. During a flight, the pulses are directly recorded on each oscillograph. They are also recorded on each tape recorder as a frequency deviation in the telemetering 40-kc frequency channel.

A voltage oscillator of reactance-tube type is used with this auxiliary equipment. Play-

back of the marker to the oscillograph is by means of a discriminator similar to those previously described. The marker discriminator, however, need not meet the rigid accuracy requirements of the data discriminator; thus, the marker discriminator is correspondingly simpler.

OBTAINING ACCURATE FREQUENCIES. A number of accurate frequencies required for proper operation of the ground-station telemetering equipment can be obtained from a commercial standard such as the American Time Products standard 2121A-2.

Available outputs are 1000 and 240 cps at high impedance. An output of 60 cps, 110 volts, 10 watts is also furnished to drive the oscillograph timing motors. The 240-cps output is used as a reference frequency on the tape recorders. The 1000-cps output is useful as a frequency standard against which to calibrate other frequencies. This circuit has been modified to improve the wave forms of the 240- and 1000-cps outputs, and a gain potentiometer has been added to control the 1000-cps output level.

A switch has also been incorporated to permit the output of 60 cps, 10 watts, to be controlled from the 240-cps interval tuning fork or from an external 240-cps source.

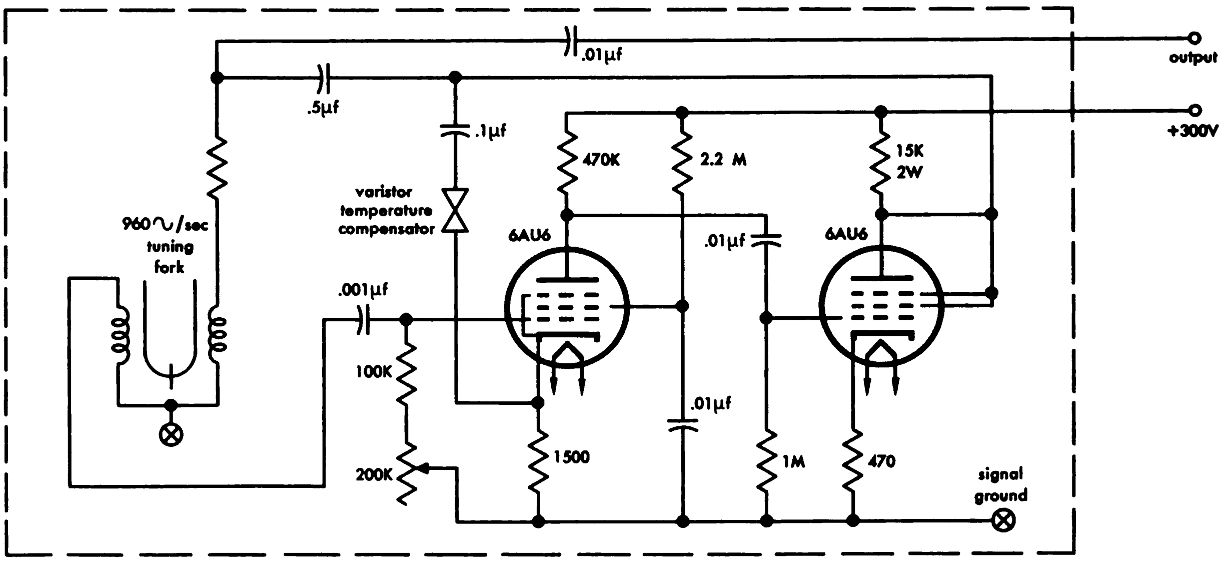
Generally, to obtain frequency standards, a tuning fork of some basic frequency, such as 240 cps or 960 cps is used. As the fork vibrates, it may cause the reluctance of an oscillator tank circuit to vary, or it may be used to actuate a multivibrator circuit or a series of frequency-multiplier or frequency-divider circuits. In place of a tuning fork, crystal-controlled oscillators with high stability characteristics may also be used as frequency standards in some applications.

An auxiliary oscillator employed in conjunction with a conventional cathode-ray oscillograph may be used to measure frequencies by means of Lissajous patterns. The 1000-cps standard frequency from the American Time Products fork is used as the reference frequency or time base.

Such an oscillator is termed an *interpolation oscillator*. In some stations, a modified Hewlett-Packard model 200 IR oscillator is used for interpolation of data frequencies as well as a source of test signals for calibration of equipment.

INTERCOMMUNICATING. For intercommunication between ground receiving stations and associated facilities, type EE7B Signal Corps telephones are widely used.

So far, in regard to internal methods of missile instrumentation, we have centered our attention on telemetering systems. In the next section, we switch over to a discussion of direct recording instrumentation, another means for collecting and recording data from experimental missiles in flight.



Basic 960 cps-oscillator unit time-base frequency standard American time products (type 2001-2)

Direct-recording Instrumentation

Collecting and recording data from guided missile can be readily accomplished by means of recording-type meters or by photographic means. Direct recording is the simplest and most accurate method for obtaining flight data, but the method is of value only when the missile or recording equipment is to be recovered.

In some recoverable research missiles, direct-reading instruments employing tapes for graphic or magnetic recordings are employed as auxiliaries to telemetering equipment. Photographic systems are likewise used. Thus, the data recorded on tape or film is used to check the data received from the telemetering system. In this way not only can characteristics of the missile be determined, but also the accuracy of the telemetering system.

For types of data subject to very rapid changes, suitable direct-recording devices have not been perfected; consequently, some conversion link must be used between the data source and the recording instrument. Frequency dividers or *down counters* are examples of such intermediate links for slowing down data to some definite proportionate rate which is within the range of the recording instrument.

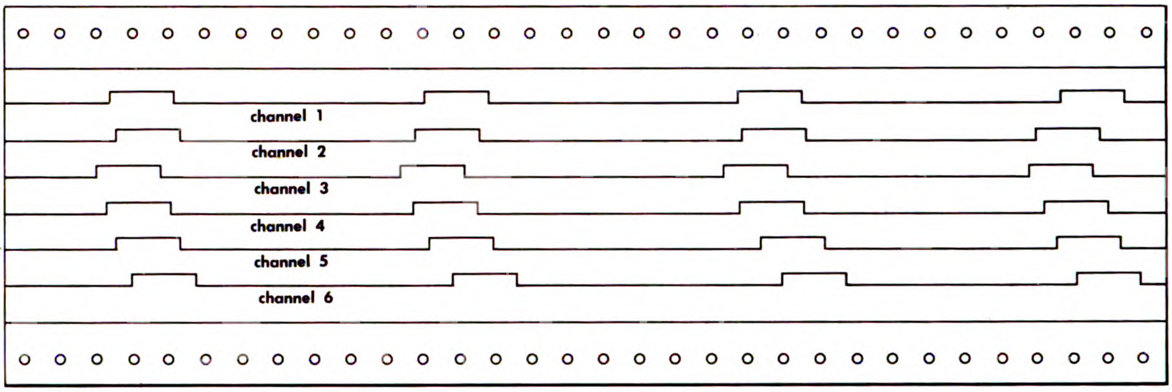
ESTERLINE-ANGUS OPERATION RECORDER

A widely used type of recorder, suitable for use in either a missile or a ground station, is the Esterline-Angus operation recorder. It consists essentially of four main components: the case, the chart drive, the writing system, and the electromagnet assembly.

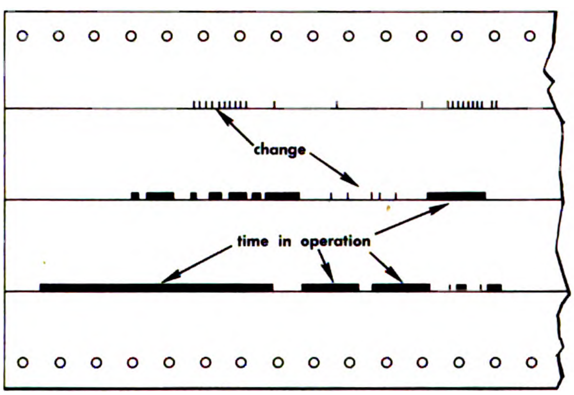
Electrically actuated, the instrument may have from 5 to 20 pens simultaneously recording the outputs of a like number of data pickoffs. Data is taken from sources whose information is of the ON-OFF, YES-NO type, indicating the time and duration of the data event and how many operations or changes occur. The recorder's charts are rectilinear. This characteristic makes it easy to compare the time of various recordings.

Esterline-Angus recorder charts are made of special grade paper, 6 inches wide and 103 feet long. They can be obtained with or without time-calibration marks along the edge. The chart of the operation recorder is propelled by a self-contained spring clock, a self-contained synchronous motor clock, an external motor or timing machine, or by a combination of these devices.

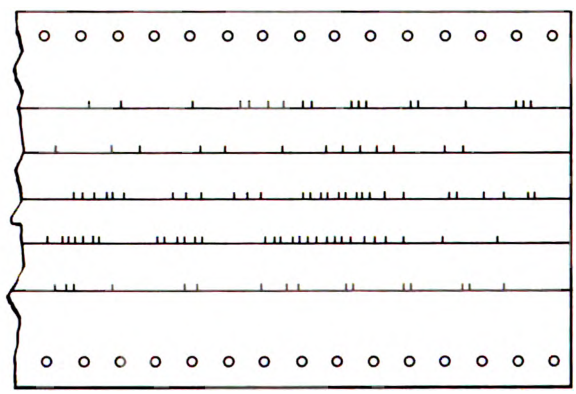
Where time intervals must be measured accurately to less than one second, a chart speed of at least ¾ inch per second is used. Speeds from ¾ inch to 3 inches per second are made possible by an external motor drive. The chart normally travels at a rate of ¾ inch, 1½ inches, or 3 inches per hour. When a disturbance occurs, the chart is automatically engaged with the external motor by solenoid clutch, immediately accelerating the clutch to 3600 times its normal speed. This high speed, in inches per second, is ample to permit accurate recording of such data as rapid circuit-breaker operations. After 24 seconds, the chart is automatically reset to the correct time of day and resumes operation at normal speed.



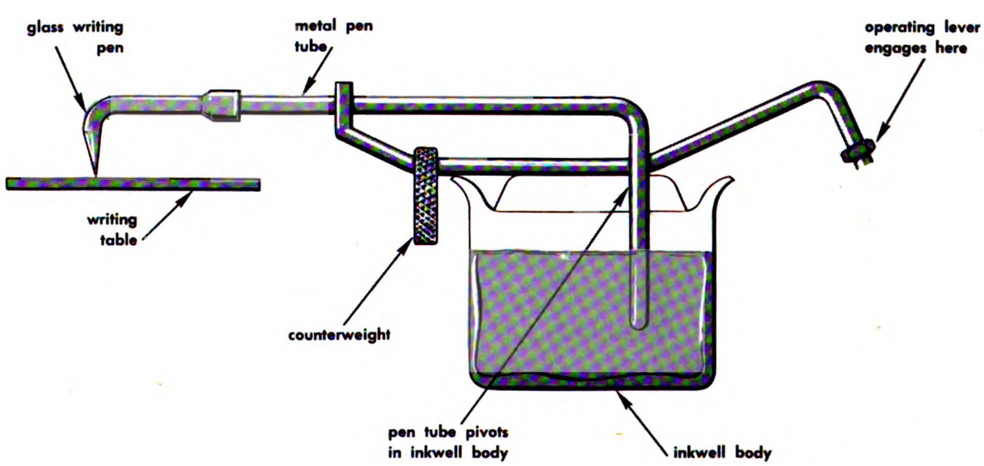
Rectangular-type record



Band-type record



Impulse-type record



Cross-section of writing mechanism of Esterline-Angus recorder

Consolidated 5-114-P4 oscillograph

Three types of record patterns are obtained on the chart, depending on the relation of the length of time the pen is energized, the time the pen is deenergized, and the speed of the chart. These types are the *rectangle*, *band*, and *impulse* patterns, so called because of their appearance on the chart. Note the accompanying diagrams.

In the Esterline-Angus operation recorder, all of the pens are served by a single inclosed inkwell, which is designed to mechanically support and align the pens. Observe in the drawing on the left how a pen tube dips into the ink supply through a hole in the metal cover of the inkwell. The small counterweight, shown fastened to the pen tube, holds the pen on the chart paper.

All pens are the same in construction and are thus interchangeable. The writing pens are made of transparent glass, friction fitted to the silver pen tubes. This construction allows for visual inspection of ink feeding and easy replacement when necessary. Divided inkwells are sometimes used to that one group of pens can record in red, another in green, etc.

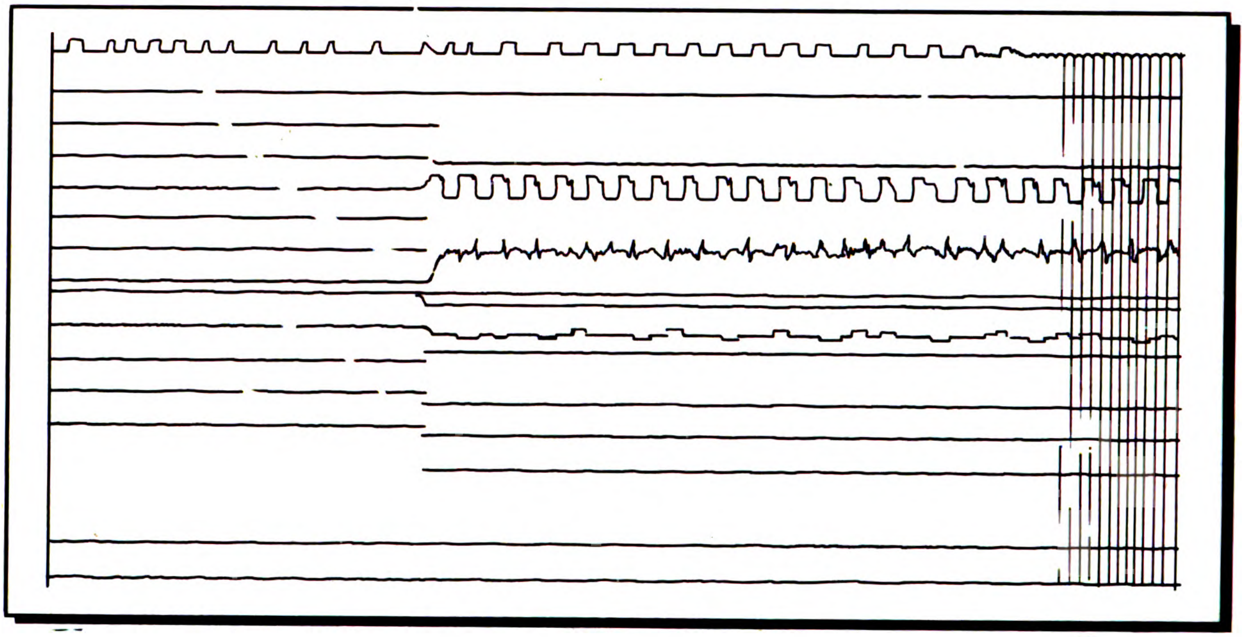
Since the drag of the pens on the chart is light, the chart driving mechanism has ample power to overcome any drag which might produce a time error. Movement of the pens is laterally across the chart, and each pen is actuated by an electromagnet which receives its energy from the pickoff source.

Each electromagnet element can follow as many as ten complete ON-OFF cycles per second, provided the ON and OFF periods are substantially equal. There must be at least 0.05 second of energized time to allow the pen to complete its stroke reliably. If the pen is normally energized, the deenergizing impulse must similarly last for at least 0.05 second. The simplest method of actuating the electromagnet elements is by the direct application of voltage.

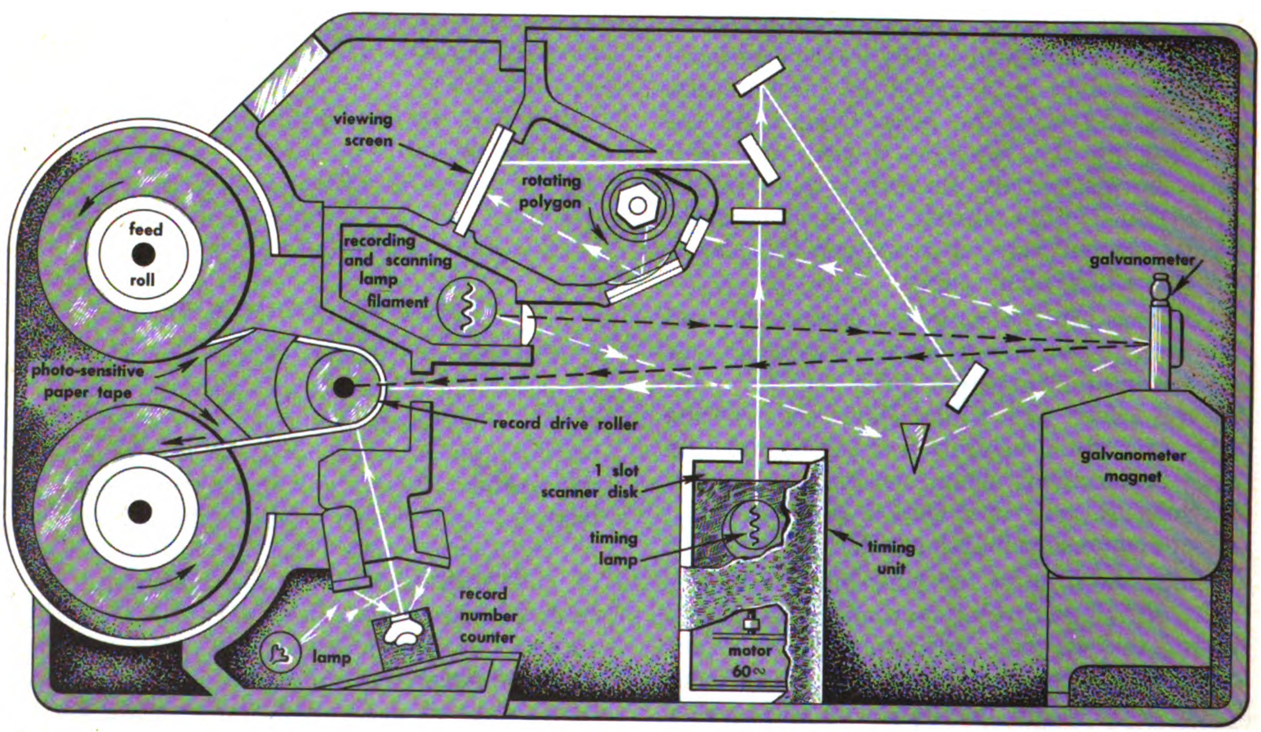
For data measurements involving changes of more than 10 cps, or very wide ranges of voltage variation, instruments of the type just described are not suitable; thus, devices such as recording oscillographs are employed.

RECORDING OSCILLOGRAPHS

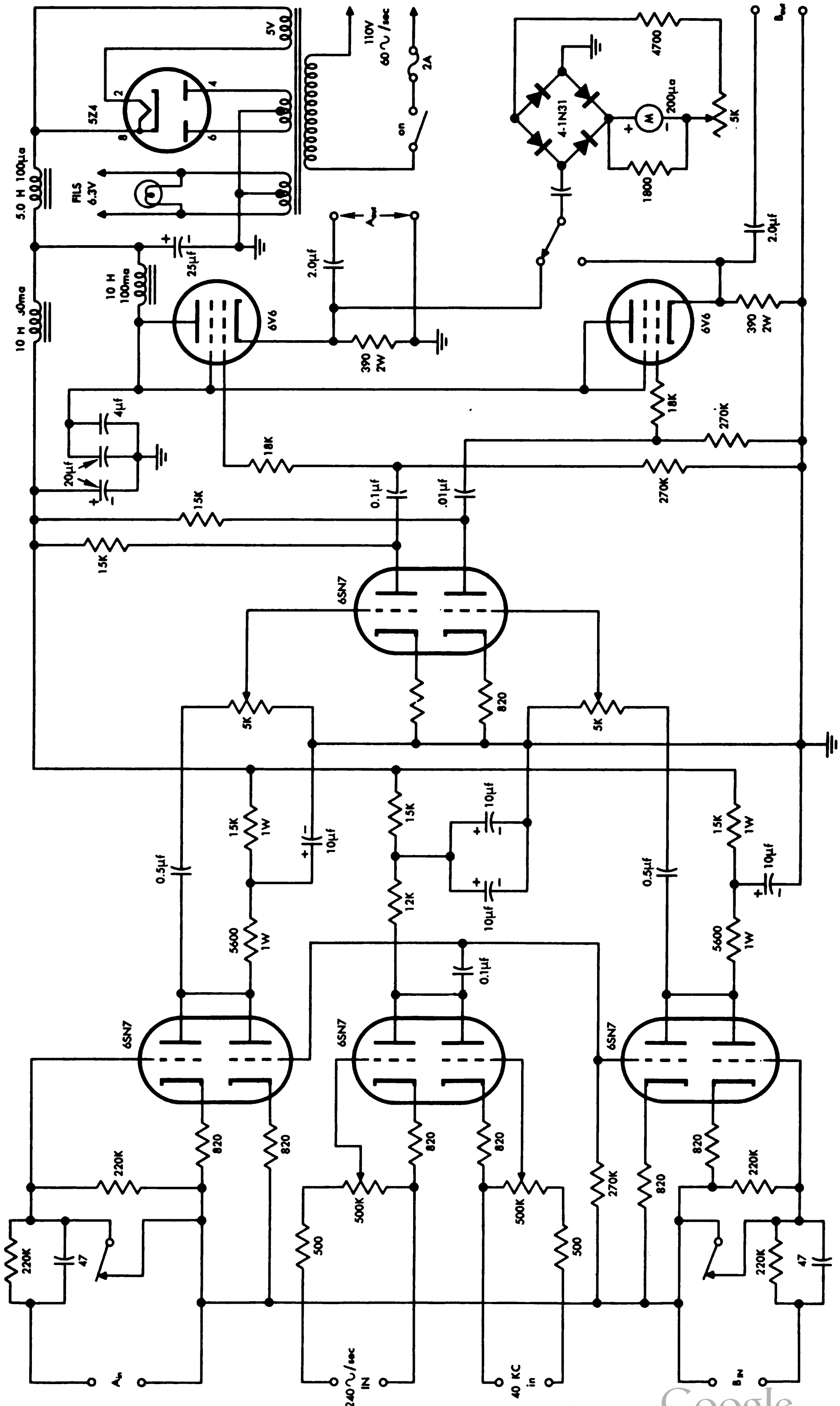
A widely used recording oscillograph is the Consolidated Engineering recording oscillograph, illustrated in the above figure. This oscillograph contains up to 18 recording galvanometers. These galvanometers deflect a beam of light across a photosensitive roll of paper in accordance with the output current of the data pickoff or the output of the receiver-data channel discriminators. Note the cutaway on the following page. The photosensitive paper is rolled past the galvanometer light beam at a constant rate, producing a permanent record on the paper, as shown on the next page.



Consolidated oscillograph record



Recording oscillograph



Wiring diagram of ampex mixer

The consolidated oscillograph is a highly flexible instrument and is well suited to recording applications in telemetering.

The time base for the recording is the speed at which the paper is run through the recorder. The paper speed is continuously variable over a wide range, and a large selection of galvanometers is available, permitting the recording of many types of data. The machine has been modified so that the timing-line motor is driven from an accurate 60-cycle/second source.

AMPEX MODEL 302 TAPE RECORDER

In conjunction with a recording oscillograph, a supplementary recording devise is used. The most commonly used type is a *tape recorder* such as the Ampex model 302.

This machine is used to record directly the output of the receivers and the calibration oscillators. The process is sometimes termed *raw-data recording.*

This type of recording possesses many advantages. If the oscillograph should fail, the tape-recorded data is still available. If a frequency response different from that originally recorded on the oscillograph is desired, the original data is available. When desired, different time bases, which call for different paper speeds, can be gotten by this process. Frequently, a large number of crossovers of galvanometers occur on the original oscillograph recording. By repositioning the galvanometers and playing the tape-recorder data back through the system, confusion of traces can be eliminated.

One of the most useful purposes of the machine is to play back to the oscillograph only selected channels, the others being removed from the record. The playback records are sufficiently accurate to permit making the data reduction of the resulting record with greater ease than from the original oscillograph record.

Accuracy of recording and playback are improved by use of an external capstan amplifier. Such an amplifier uses a stable 60-cps fork as a standard and amplifies the voltage to 110 volts, 75 watts. This insures constant tape speed (uniform time base) when the equipment is operated from a gasoline-driven generator with poor frequency stability. It also improves results when the equipment is operated from commercial 60-cps power, which, although accurate over long average periods of time, is subject to considerable variation and fluctuations for short time intervals.

A mixer amplifier, such as the Ampex mixer in the wiring diagram on the preceding page, supplements the tape recorder and raises the input signals to a good recording level. Note that 40-kc/sec time pulses and a 240-cps sine-wave reference tone are also mixed with the signals. The reference tone is accurate according to the American Time Products frequency standard.

Since an accurate 240 cps is recorded on the tape, variations in tape speed from any cause either in recording or in playback can be detected. These variations can be corrected by comparing the frequency of the tone on playback against the same tone generated by the fork at the time of playback. Accurate time lines can also be recorded from the tape to the oscillograph record by feeding the recorded 240 cps through the frequency standard and converting it to the 60-cps outputs seen on the right side of the diagram. This output is used to drive the oscillograph time-line motor, or it is used to drive the operation-recorder-chart drive motor.

As stated earlier, direct recording is the simplest and most accurate method for obtaining flight data. But the method is limited in use in the missile field because the recording equipment or the missile containing the equipment must be recovered to be of value. When used, instruments such as pen recorders, oscillographs, tape recorders, and cameras are highly valuable to missile recearch in direct-recording systems.

Guided Missile Short-range Guidance System

Most of this supplement is concerned with command guidance. However, before taking up command guidance, let's briefly consider preset guidance systems.

PRESET GUIDANCE SYSTEMS

A self-contained system of guidance falls into the category of a preset system. The intelligence necessary to the course and termination of the flight is all set into the missile prior to its launching. However, the usual definition of preset guidance includes only a simple system that defines heading, altitude, time or length of flight, and programmed turns.

A simple preset system can be broken into two functions. One function is to determine the heading and the other is to determine the distance of flight. This portion of the explanation is therefore divided into two corresponding sections. Various means of accomplishing each of the functions are considered.

The most obvious form of *heading* reference is that form utilizing the earth's magnetic field. For many years the magnetic compass and its refinements have served as a heading reference. A control arrangement, making use of a gyrosyn compass, is an example of a preset reference for heading. The yaw gyro of a control system is another typical example of a preset reference. If a yaw gyro is not capable of being precessed from an external source while in flight, the heading set in prior to launch controls the path of the missile.

The length of flight can also be taken care of in a simple fashion. The distance a missile has flown can be determined by the use of an air log or by summing (integrating) the average velocity of flight for the time the missile has flown. An air log consists of a small propeller of known pitch, turning a Veeder-root counter. Through a precision gear box connecting the propeller and counter, the counter can be made to show miles, feet, kilometers, or an arbitrary measure of distance.

Air pressure developed by missile movement can be used to make an airspeed meter give

an indication of velocity. By integrating (summing the average) airspeed over the time of the flight, a continuous check of the distance flown is obtained. This method has the same accuracy as the air log, and the only reason for choosing one or the other is the availability of components.

The latter method could make use of standard equipment but would usually require an electronic installation. The air log equipment would be all mechanical, merely energizing a switch or relay at the dump point.

Another possible method of determining distance flown involves an airspeed reference set into the controls section and a clock for time elapsed in flight. This method is the most inaccurate. Besides being subject to all the error-creating influences that the other systems are subject to, the reference system's accuracy is affected by the control capabilities of the missile to maintain closely the airspeed reference.

The simplest form of a preset system, as used in almost all missiles, is known as launch bias. Because of the tremendous acceleration occurring at the launching of a missile, the mechanical and electromechanical components do not function properly. To take the missile through this initial stage of its flight, fixed control and throttle setting are made before the launching takes place. These control settings are calculated for a stable climb-out during the initial acceleration, until the regular giudance and controls can take over.

In addition to heading and distance, a third aspect of flight exists. This aspect consists of altitude reference. An altitude transducer is used to control the heat of the missile flight. An altitude transducer is an altimeter with an electric output. The simplicity and accuracy of the altitude transducer have resulted in its remaining as the primary altitude control or reference in even the latest guidance systems.

The preset system was the earliest of the methods of guidance, and in some form will undoubtedly be the longest in use. The settings for launching missiles are always necessary whether they are made manually as with the surface-to-surface missiles, or automatically as set in by gunsight computers in air-launched missiles; whether they last for many minutes as in the World War II missiles, or mere seconds as in preset systems.

COMMAND GUIDANCE SYSTEMS

A command guidance system is one by which the missile receives correction signals from an external source. Some enlargement on the difference between correction and error signals becomes necessary at this point. A correction signal is a command that activates the controls a definite amount. An error signal is the detected discrepancy from the required course, altitude, or speed. Ordinarily the guidance system detects and develops the position error signal into a correction signal. After the correct signal is properly formed, it is fed into the control system.

Information Link

An important link in a complete command system is the method of obtaining information as to error in the missile's position. Some means of distinguishing the desired position from the actual position, and the amount of error existing between them, must be available so that the command operator can instigate the proper signals. This *information link* can be *visual* or *electronic*. The electronic means can include television, radar, and certain forms of hyperbolic systems. Methods of returning information from missiles are covered under the classification of *instrumentation;* considerable detail on these methods can be found in later chapters. The *visual link* utilized telescopes or the unaided eye.

The accuracy and dependability of the information link actually determines the accuracy and dependability of the complete command system. Much effort is made to accomplish an accurate, dependable system for obtaining information from the missile. A system limited to daylight use is the television relay, in which a TV camera is located in the nose of a missile. The camera's video signal is transmitted back to the controlling point.

A quite similar arrangement, which has no visibility limitations, uses a radar set in the nose of the missile. The set transmits the video signal and antenna orientation signal back to the control point. Other missile information

links are the systems which are located at the control point and which track the missile.

Command Links

The equipment used to send command signals to a missile actually performs the function of a communications link. If a human pilot were in the missile, voice radio would be used. However, in the case of missiles, the commands must be sent in the language of the missile guidance and control equipment.

Command systems are the actuality of the popular conception of remote control. In fact, whether simple or complex, a manual control unit can only operate with the accuracy of the human manipulating it. For increased precision the human operator is replaced by an automatic device. The output of the system tracking the missile is fed into a computer. The computer also has the information as to the location of the target set in, and automatically calculates and keys the command transmitter in the proper sequence to effect a hit.

The command phase of remote guidance has been under development for many years. Its principles are quite well known. The target drones and other such familiar equipment are examples of command guidance.

One of the later developments in command guidance systems is the utilization of a combination command and tracking radar in the capacity of a command link. A beacon receiver is used in the missile as the command receiver. This development has a distinct advantage over previous equipment, since there is one less electronic link to have malfunctions or be subject to interference. Also, the amount of equipment and consequent power requirements are lessened. The principles of this combined tracking-command system are covered later.

To illustrate the *principle* of command guidance, let's first discuss the early concepts and then make a short survey of the technical evolution leading to present-day systems and principles.

There have been many approaches, some quite ingenious, to the problem of command guidance. Perhaps the most straightforward of the early approaches was of a *multiple transmitter and receiver* installation. In this arrangement the controlled craft has a *separate receiver* (usually one tube) for *each command* function to be performed. The transmitter was either multiple or capable of transmitting its carrier on any of the several radio frequencies for which a receiver was tuned. This system had the advantage in that no channel interaction was experienced, but it had the disadvantage of being bulky and requiring too wide an RF spectrum. Because of the number of receivers, only the most simple and therefore insensitive receivers could be used.

The natural refinement to this type of system was the use of *tone channels*. By modulating the transmitter with various audio frequencies, it is possible to use one RF carrier and receiver. After demodulation of the carrier in the receiver, the audio frequencies are used to excite the tone channel, which is resonant to the particular frequency for the function they perform. In the all-RF system, the number of functions capable of being accomplished was limited to about four, due to space and channel requirements. By using a tone-modulated system, ten audio channels can be used without undue complications. Then by proper combinations of the audio channels, 18 or more functions can be performed. In new "miniaturized" equipment, 20 primary channels are to be available, giving a possibility of many more functions.

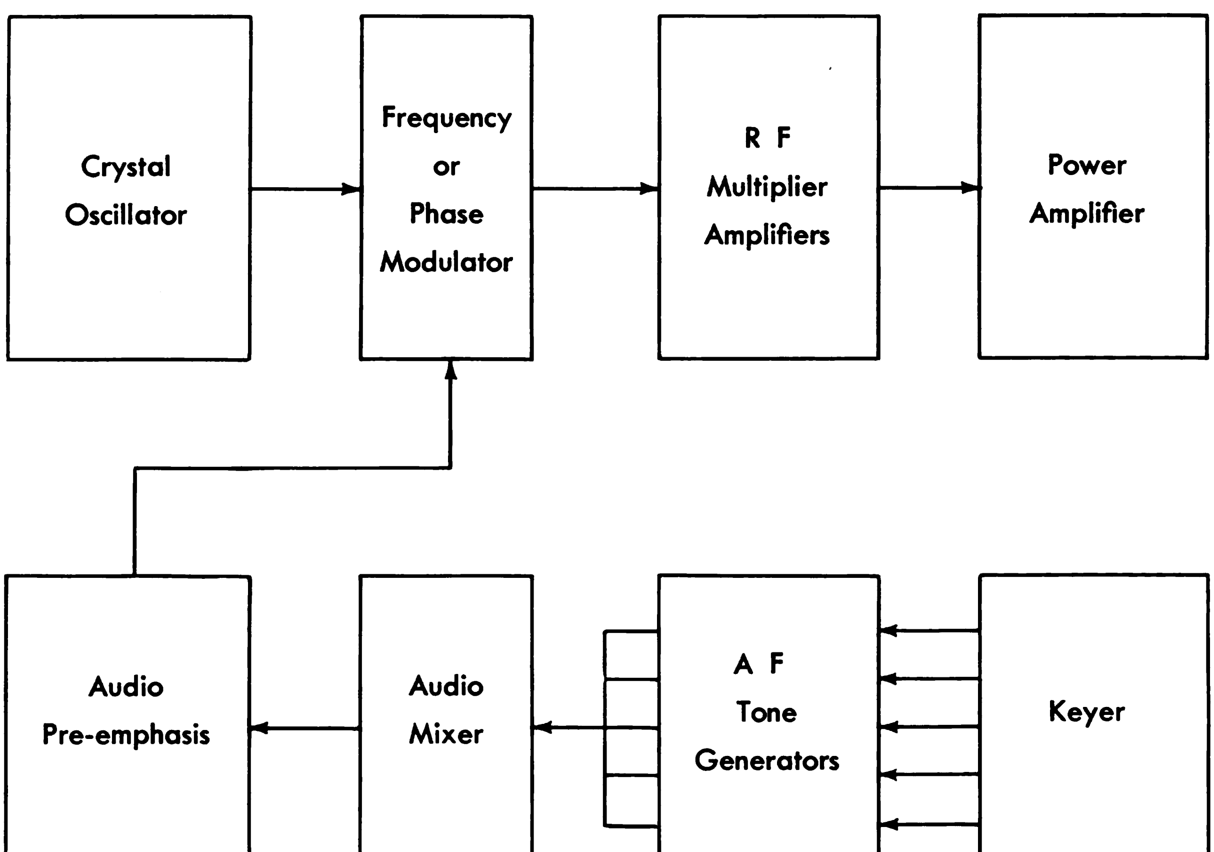
The tone-modulated system developed new interference problems which became apparent especially in the running of development tests where accuracy is a necessity. Interference from other sources containing the frequency of a tone used for a control function causes a craft to react. Often a harmonic or sideband of a voice-modulated signal contained frequency components sufficient to upset a whole operation. The change to frequency modulated systems eliminated much of this interference. However, man-made interference of an FM nature still could create difficulties. To overcome this, coding combinations of the tone channels were devised. Certain operations could not take place in the craft unless several selected tones appeared

at the receiver simultaneously. The chance that an interfering source could duplicate this combination was negligible.

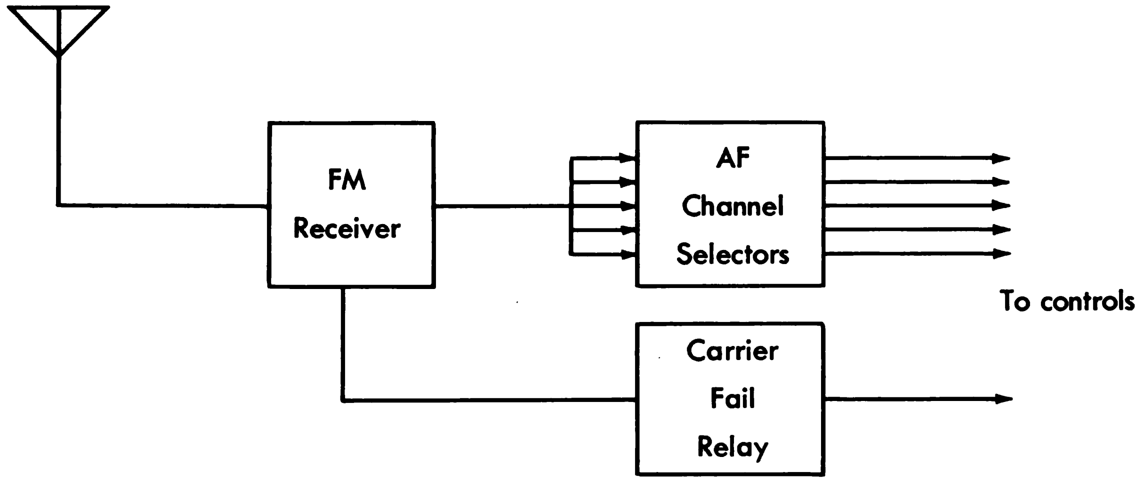
COMMAND TRANSMITTERS. The RF channel of the command transmitter is not unlike any other FM or PM transmitter, as evidenced by the accompanying diagram. A quartz crystal is used for stability and accuracy of the carrier output. The modulation of the carrier must necessarily follow the oscillator if stability is to be retained. After modulation, the necessary stages of RF multiplication raise the carrier frequency to an output in the VHF band. The power amplifier furnishes the RF power necessary to send information through space to the receiver.

Notice in the accompanying diagram that in the lower audio channel the peculiarity of command equipment is evident. A group of audio tone generators .are present. Each generator operates only when it is actuated by the circuits in the keyer, which may be manually controlled or computer controlled. The individual outputs of the tone generators are mixed together into a composite audio signal and applied through the audio preemphasis network to the modulator.

The preemphasis network is included in order to maintain the audio signal-to-noise ratio at optimum operating conditions for the overall system (including the receiver). The function of this netowrk is to emphasize the higher audio frequency tones which are later deemphasized in a network of opposite characteristics in the receiver. This operation causes the signal-to-noise ratio to remain more constant throughout the audio range. Because the noise appearing with the signal consists of high frequency components, it is only necessary to perform the pre- and deemphasis on the audio frequencies in this range.



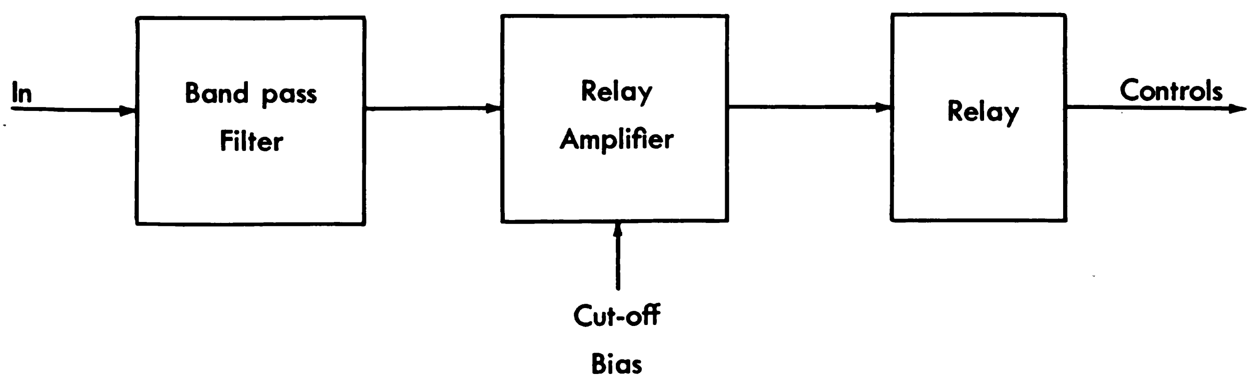
Command Transmitter



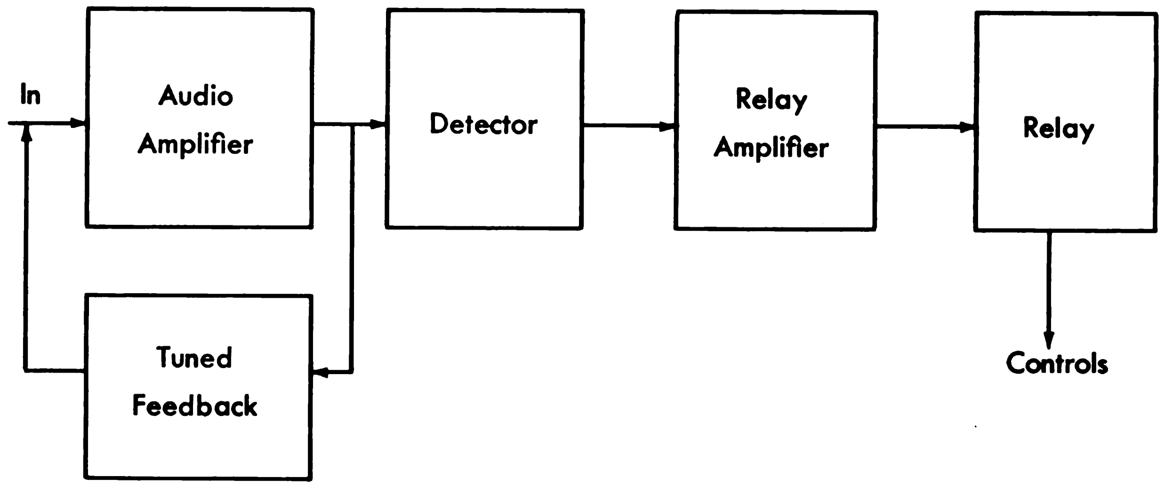
Command Receiver

COMMAND RECEIVER. The circuits of the receiver shown in the above diagram are those of a standard FM communications receiver with some added refinements. Ordinary superhetrodyne connections are used through the second detector. At the limiter stage preceding the detector, the limiter grid current, which is proportional to the strength of the incoming carrier, is used to operate a carrier fail relay in case the carrier should get too weak. The transmitter carrier is usually left on, even if no modulation is being sent. It is left on so that the receiver will have a means of determining whether it is continuing to receive a signal.

After amplification, the discriminator (detector) output is applied to the selector channels. It is these selector channels and their operation that make the receiver unique in its application. There is one selector channel for every tone that the transmitter may send. The selector channel, a breakdown of which is presented in the diagram below, is usually an amplifier with a band-pass filter at the input and a relay in the place circuit. The selector tube is held below cut-off bias until the tone to which the input filter is tuned appears. This tone is passed by the filter and appears on the grid, causing plate current to flow. The plate current flowing in the coil of the relay energizes the relay, completing the circuit for performing the desired function.



Selector Channel Using Band-Pass Filter



Selector Channel Using Tuned Feedback

The diagram above shows a variation of the preceding selector channel. The tuned feedback is adjusted to aid the audio amplifier in amplifying only the desired frequency signals. Detection of the signal is then made before applying it to the relay amplifier.

A rather unique combination for issuing command signals is the use of tracking radar and command receivers. In order to extend tracking ranges on the relatively small missiles which offer poor radar return, a radar beacon is installed in the missile. This radar beacon transmits an echo pulse that is much stronger than the echo that would be returned on *skin track* (without a beacon).

The radar beacon is a small receiver and transmitter operating in the tracking radar band. The receiver has circuitry that only accepts a radar signal that has a definite pulse separation. In other words, it only responds to a certain coded interrogation. If the proper pulse combination is received, the beacon transmitter sends out a pulse of RF.

The transmitter operates on a slightly different frequency than the exact frequency of the radar. The radar receiver is tuned to this slightly different frequency so it can detect the beacon signal without having it confused by the ordinary return of the radar transmitter.

The beacon receiver can also be arranged to accept command signals. An additional set of coded pulses is added to the transmitted signal from the radar, arranged to give some intelligence such as a command tone. The beacon set accepts these signals and channels them into a different function, which decodes them for the intelligence they contain.

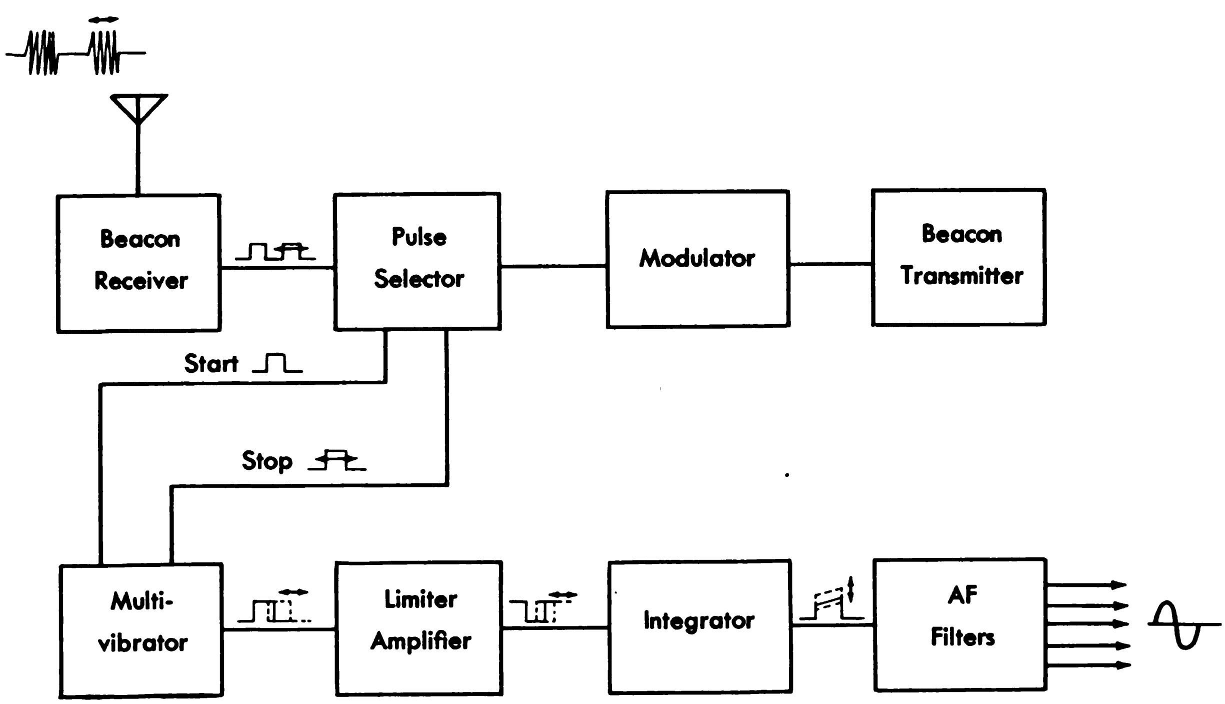
This intelligence is impressed on the radar signal as pulse position modulation. This modulation gives the autopilot its commands just the same as a human controller voices his commands to a pilot over the radio communications circuit. Pulse position modulation requires but two pulses per sampling cycle to give one intelligence channel. It is possible to include other intelligence channels in one sampling cycle by tone modulation, as was explained earlier. The beacon interrogation pulses are also in the pulse train received from the radar. The position of the modulated pulse is varied back and forth about its rest position at an audio rate. The number of intelligence channels is limited because of the limitation of audio modulation which can be accomplished by the sampling frequency and because of the fact that the radar PRF (which is the sampling rate) is dictated by the maximum required range of the radar set.

Now, examine the diagram on the following page. Note how it is possible to use the beacon as a command receiver.

Let's trace the path of a signal through the diagram to show the transition from the pulsed RF input to the sine wave audio output. The receiver converts the RF signal to a set of video pulses which are separated by the pulse selector. The first pulse is used to start, and the second modulated pulse to stop, the multivibrator, thus producing a square wave. This gives a square wave of a varying width. The varying width follows the excursions of the modulated pulse. An integrator forms a sawtooth wave whose amplitude varies with the width of the square wave. Thus the pulse position or time difference in the received signal is converted to the amplitude difference of a sawtooth wave, whose frequency is the same as the radar PRF. The sawtooth waves of varying amplitude are fed into a frequency selective amplifier, and all components except the fundamental modulating frequency (which is evidenced by the amplitude variations) are removed. The resulting sine wave output then excites the relay amplifier and institutes the desired control function.

The criterion for selection of tones to be used in any command system requires that there be no possibility of a false signal being created through interaction of other signals. The audio tones must be kept below a certain frequency to limit the sidebands. It must not be possible for any harmonic of any tone, or beat note of any combination of tones, to be able to actuate a receiver channel which is not desired. The practical bandpass characteristics of the selective circuits make it difficult for them to be high "Q"; therefore, the separation between tones must be even greater than if high "Q" circuits could be used. A set of frequencies to satisfy the above conditions, therefore, requires quite a bit of study.

Inasmuch as the commands are so closely tied in with the control section, it is usually the control capabilities or requirements that dictate the method of controlling the modulation of the command link.



Combination Radar Beacon and Command Receiver

In devising the control unit for keying the transmitter, the ultimate in development is proportional control. This means that a signal is sent which is just large enough to correct the particular error and no more.

A contrasting method of control is an on-off system, which means full-strength control applied for the period of the command signal. An example of the on-off system is control by a stepper box. The stepper box is a control box whose main component is a four-way toggle switch mounted so that its movement simulates that of an aircraft control stick. The operator merely holds the stick in the proper position to bring about the desired change.

It is not too difficult to realize what the normal outcome of this type of control could be. A pilot controller commands a function until it is evident to him that it is taking place. At this time the controls are well advanced so that there is a greater movement of the craft than desired. A countercommand has to be sent to correct this overcontrol. The missile then tends to follow an oscillating course which, if not properly damped, could become uncontrollable. There are other switches located on the box for various control functions.

The closest approach to remote *proportional control* is a method by which the controller is varied in small increments to obtain smooth action and to allow the aircraft to receive the control necessary to produce an exact amount of change. The equipment in the missile follows these incremental moves exactly. This is done with a pulsing arrangement, much like the selector actuated by a telephone dial. The command operator elects the number of pulses to be sent. Each pulse produces an incremental change in the desired missile control unit. The unit would be a selsyn or some other type of pickoff. The pulse repetition rate is determined by the pulsing equipment, just as it is in the free rotation of a telephone dial.

In order not to lose the reference position, the operator may automatically step the missile unit to the position by pressing an index button. This indexing arrangement checks for lost or extra pulses. The index usually requires the use of another channel in the command equipment. Actuation of the indexing channel causes the selector in the missile to return to a reference point.

Because of the close tie-in that command guidance has with the control section, some manufacturers consider command guidance as part of the autopilot development. Command guidance can be compared to an extension of the control cable link between a pilot's hand and the control surface of a piloted aircraft.

Short-Range Hyperbolic Navigation

The range of hyperbolic navigation depends upon the radio frequency that is used as the carrier. Here we will discuss hyperbolic navigation as it applies to short-range guidance systems.

If it is desired to have a short-range navigation system of good accuracy, hyperbolic principles can be applied to accomplish this. By utilizing microwaves, a small, highly directive antenna can be contained in the missile without interfering with its aerodynamic characteristics. The main problem is to eliminate errors in the received signal by discriminating against sky waves. The directional characteristics of the antenna are narrow in the vertical plane and fairly wide in the horizontal plane. An important effect of the directional receiver antenna is that it has gain — it actually does the same job as an RF amplifier connected to a standard dipole antenna. Also, its directional characteristic decreases the possibility of countermeasures by jamming. Mose of the sky wave discrimination exists because transmitted energy can be compressed in the vertical pattern, and the line-of-sight transmission to the missile affords little opportunity for interference from multi-path signals.

This discrimination against sky-wave interference is necessary in the synchronization of the ground stations. The synchronizing pulse must be transmitted via a direct, constant path to be accurate. There should be no variable factors such as *skip effect* which would alter the transmission of synchronizing signals. Establishing a condition in which

these variable factors do not alter the transmission poses a greater problem with UHF systems using a base line longer than the line-of-sight distance. A means which does not introduce any varying or unpredictable delays is used to relay the synchronizing signal. The transmitter stations have precision timing signal generators whose outputs are used to modulate RF transmitters that are typical for the band employed. That is, they use the same sort of tubes and circuitry as would a radar transmitter in the same band. The slave ground station also requires for the synchronizing signals a receiver that is quite similar to the reception equipment utilized in the missile. The transmitters must have high power output to give a high signal-to-noise ratio at long ranges. The separation between ground stations is normally limited to less than 100 miles because of the limitation of line-of-sight distances. This short base line would result in a very short-range system if a three-station (one master and two slaves) system were used. The lines of position (LOP) in such a set-up would cross an extremely obtuse angle and would make an accurate fix difficult for any point located at any distance from the base line.

A four-station system has the advantages of both the line-of-sight transmissions and long base line systems. Two pairs of stations are used; each pair consists of one master and one slave operating on the same frequency, and they are properly synchronized. The pairs of stations are separated by some distance so that the resulting hyperbolic grid system has the LOP more nearly normal to one another in the desired target areas. The illustration on page 407 of the unclassified text shows a *two-station* hyperbolic grid. You can see that if a second similar system is uperimposed on the two-station grid at nearly right angles, the target points would be more positively located due to the LOP crossing at angles closer to a right angle.

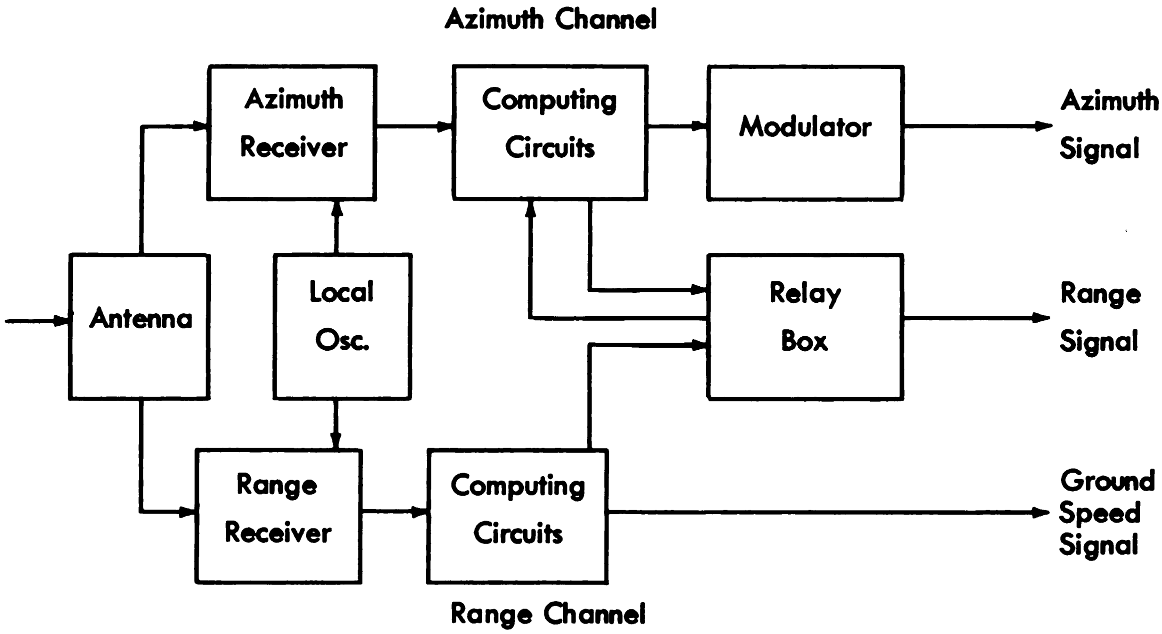
One pair of guidance base stations is used to give the azimuth guidance hyperbola. One time difference line of this pair (a hyperbola) is chosen that will cross the target area and thus serve as the desired track of the missile. The guidance equipment within the missile can determine when the received signal pulses have the proper *time separation* and show *on course.* If the received guidance signals do not have the desired time difference, the guidance equipment can determine whether the difference is greater or less than the on-course condition, thus determining whether the missile is right or left of the desired course. The error signal from the guidance section is then sent to the control system for processing into the proper control correction in yaw.

The other pair of guidance base stations is used for range determination. A particular time difference line of this system is calculated to pass through the correct point for instituting the terminal phase of the flight. The intersection of this *range line* and the *course line* gives a fix at the predump point. The missile guidance equipment develops a voltage output proportional to the distance from the preselected target point. This varying output voltage approaches zero at a rate proportional to the velocity of the missile.

This output voltage is converted to a voltage proportional to the rate of change of the output voltage. This proportional voltage is then a measure of the missile ground speed. Circuits are necessarily incorporated to correct for the hyperbolic expansion. That is, adjacent lines become further apart as the distance from the base stations increases. The ground speed signal is necessary in the terminal phase to alter the dive path so that the missile will not overshoot or undershoot its target due to airspeed error from the preselected speed.

The output of the range portion of the guidance equipment decreases to zero volts, at which time the terminal guidance takes over. The missile then makes an approach, contacting the target area under terminal guidance.

The diagram on the next page illustrates the operation of the guidance equipment which might be located in a missile. The signals are received at the missile by a directional microwave antenna which is located so as to receive signals from the rear. A cover is placed over the end of the antenna, designed to hold air pressure around the antenna and transmitter at sea-level pressure no matter



Short-Range Hyperbolic Missile Guidance

what the altitude of flight. The cover also prevents entry of moisture.

The two receivers form the beginning of the range and azimuth channels, as shown in the above figure. The single local oscillator frequency is midway between the range and azimuth carrier frequencies. The resultant IF is then the same for each channel. The mixer is a crystal, located in the tuned-cavity coaxial connectors coupling the mixer output to the respective IF amplifier.

The IF amplifiers are stagger tuned to provide a band wide enough to preserve the shape of the short signal pulses. The output of the IF stages is demodulated by a crystal detector. The video section which follows then brings the signal level up to that required by the computing circuits.

The computing circuits first identify the master and slave pulses and then compare the time difference between them to the desired time difference. In the azimuth channel, the desired time difference represents an on-course condition. In the range channel the desired time difference represents the beginning of the terminal dive phase.

Identification of the pulses is necessary for two reasons. First, it is necessary to distinguish the slave signal from the master signal so that the proper signal becomes the reference. Second, identification is necessary to prevent false signals caused by noise or enemy jamming from controlling the equipment. Each signal must possess some peculiar characteristics in order that they may be positively identified. This is done by sending a pulse train of two or more pulses in place of a single pulse from each transmitter. The pulses may be coded by using certain amplitude differences, pulse width, pulse separation, number of pulses, or some combination of these methods.

When time difference coding is used, the two slave and the two master signals contain two pulses each. In either channel, the slave signal has a difference pulse separation from the master signal. Two decoders are used in each channel to identify the master and the slave signals.

The output of the range channel computer is used to determine the ground speed. The computing circuits are constantly measuring the distance from the missile's location to a preselected (zero) hyperbola near the target area. The output of the computer is a constant DC voltage until the missile comes within a given range in the range hyperbolic grid. At this given range the DC varies in propor-

tion to the range from the preselected zero hyperbola. This contantly decreasing voltage can now be differentiated to indicate the rate of change of range time differences (crossing of range hyperbolas) as the zero hyperbola is approached. Thus the operation becomes missile veolocity.

The changing DC output of the computing circuits is also applied to another circuit where the signal amplitude at any given moment is an indication of the range time difference (range-to-go). The phase of the wave denotes approach to the zero hyperbola or passage beyond this point. At a preset voltage level following passing of the zero hyperbola, a relay is activated, causing the control system of the missile to take over. The controls then determine the *dump* or dive point after precise automatic corrections of the range-to-go voltage.

A single site system, which does not use the hyperbolic principle, has been proposed. In such a system, two base station transmitters would be placed about a mile apart. The missile trajectory would be at right angles to the base line, the base line being a straight line between both stations. The missile would contain equipment for retransmitting a microwave signal which has been excited by the base station equipment. After launching, the retransmitted signal would be used at the ground station to determine missile position and velocity. The position would be determined by narrow-beam tracking antennas. The range of the missile would be measured by comparing the phase of the returned signal with the transmitted signal. The velocity of the missile would be determined by differentiating the range signal or by using Doppler effect.

All the position and measurement outputs would be fed into a computer to be compared with the desired course. The ground computer would calculate the amount and direction of any difference and convert it to a command signal to be transmitted to the missile for correcting its course.

The transmitted command signal may not directly affect the control surfaces. The error signals could be added to the signal in the missile inertial guidance computer in such a way as to correct the computer output.

Beam Rider Guidance

Beam rider guidance has its primary use in air-to-air or surface-to-air missiles. It requires that the launching vehicle and guidance radar remain fixed on the target until the interception is complete.

The beam rider was one of the first of the interception or antiaircraft automatic guidance systems to be suggested. Tracking radars for gun directing systems were developed and were successful. However, evasive action by a target craft after a shot was fired would result in a miss even though the radar could easily keep a track on the target. If it were possible to have a projectile that would remain with the beam, a hit would be guaranteed. The idea of the beam-rider missile was then conceived.

Beam-rider guidance is accomplished by devising a system in which the missile can sense its position in the radar beam and correct itself to remain in the center of the beam. The original plan was to have the ground radar, which was tracking the target aircraft, control the interceptor missile. Because this radar beam was moving continuously, the missile became subject to sideward accelerations which tended to strain its aerodynamic capabilities.

A two-radar system was then devised, consisting of a tracking radar and a control radar, The output of the tracking radar is fed into a director computer which pointed the control radar at a predicted position of the target aircraft. This predicted position is continuously corrected, but beam movements are very small as compared to those of the tracking radar. The missile is launched into the beam of the control radar and guided to its target. The control radar thus tracks the missile and furnishes control signals to maintain its trajectory to the point of impact. A two-radar system can obviously handle only one missile at a time, so it does not permit a high rate of fire.

A one-radar system is used in the air-to-air version of the beam rider. This means that the launcher aircraft, missile, and target aircraft have to remain in line at all times. However, it is possible to fire missiles in sequence, with a

half second or so separation, and thus control more than one missile at a time.

Consider a missile which is intended for use with an associated pulse modulated radar system. Detection of the target in such a system is accomplished by directing a beam of pulsed high frequency radio energy in a preset pattern over the space to be searched. When the beam strikes an object, energy is reflected, and a small portion is returned to and detected by the radar system. Since the beam is narrow, the direction of the target is known. And since the energy is pulsed, the elapsed time between the transmission of the pulse and the receipt of its echo gives a measure of the range.

The projected beam is not fixed with respect to the axis of the antenna reflector but is caused to trace a cone in space by nutation or rotation of the central antenna feed. *Rotation* is the spinning of the antenna dipole, slightly offset from the reflector focal point, to produce a conical pattern. *Nutation* moves the antenna in circular position about the focal point without varying its polarization; that is, the antenna remains in a horizontal or vertical plane throughout its cycle of nutation. Nutation of radar antennas is a much more efficient method of conical scan because the returned signal always has the same polarization as the antenna. In the rotating feed, the polarization of the returned signal always slightly lags the antenna position and thus gives a weaker signal than if the antenna were properly polarized.

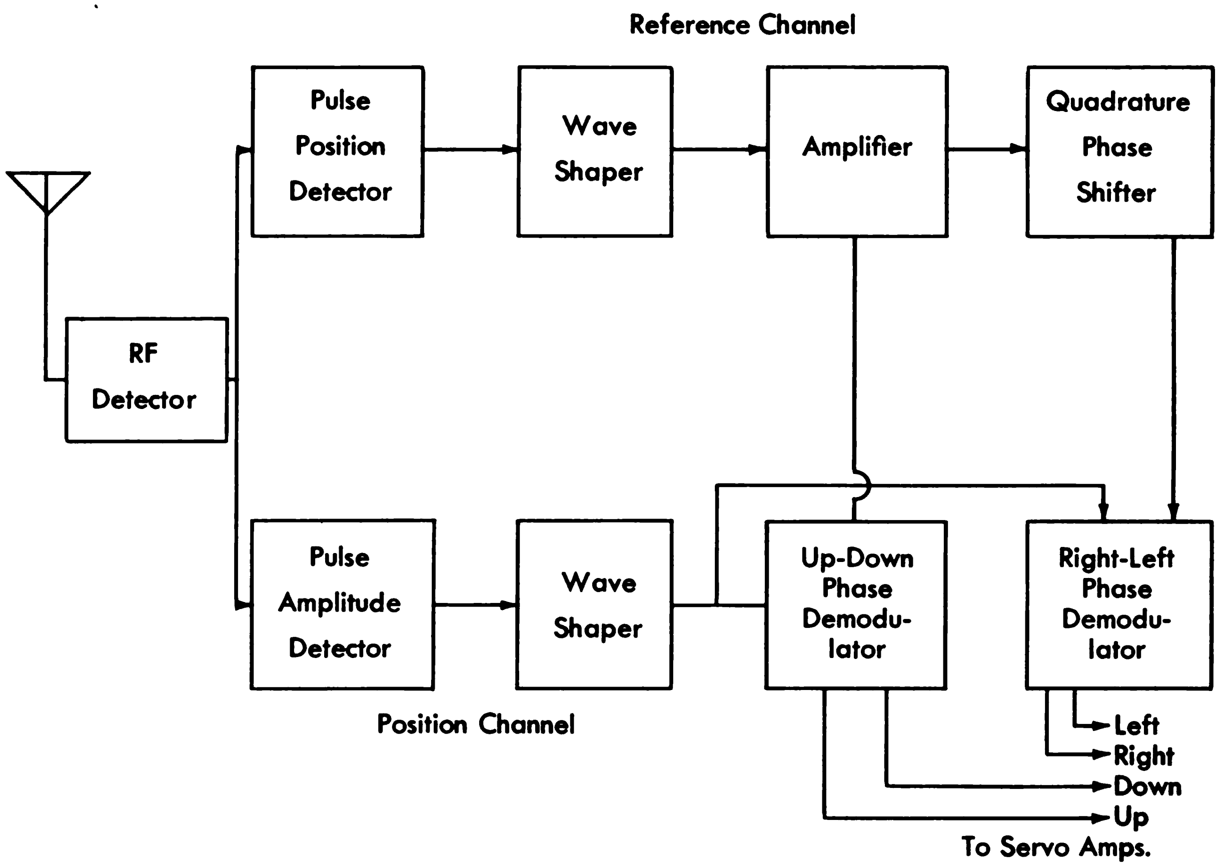
A target on the axis of the reflector always has the same echo amplitude for all positions of the beam. If the target moves away from the beam axis, the echo signal varies approximately sinusoidally with the rotation of the beam. The reflected signal can thus identify the target displacement from the reflector axis in direction. For small target deviations, the error signal amplitude indicates the amount of displacement. The axis of the conical trace of the radar beam tracking the target provides a line-of-sight route. The problem concerns the method of controlling the missile to fly along an automatically maintained path or to *ride the beam*. In order to accomplish this, it is necessary that the missile be able to sense when it is on the beam axis. Further, if displaced, the missile must know the direction and degree of displacement.

The missile has no internal phase reference as does the radar tracking system. The modulated signal of the beam must furnish the necessary guidance information. The beam is modulated in such a way that, in addition to other functions, it can either furnish the missile with the *sensing information* in a *one* radar system or can *command* the missile in a *two* radar system. The basic idea of radar command control has already been covered.

In a single radar system the spinning beam of the director radar impresses signals on the beam to identify the position of the missile about the beam's axis. These identifying signals may be in the form of coded pulses transmitted while the director radar antenna is in a specific quadrant. This coding could be in the form of pulse time modulation applied to every other tracking pulse and used in a system as shown in the diagram on the following page. This pulse time modulation, when detected in the pulse position detector of the missile receiver, would direct the error sine wave to the position channel at the conical spin or scan frequency, which is usually about 30 to 60 cycles per second.

This reference frequency is compared in the position channel phase demodulator to the sine wave resulting from the amplitude modulation on the received pulses. The reference frequency is then phase detected to give the direction of error. Therefore, correction is obtained to return the missile to the center of the beam in the up-down axis and right-left axis.

In a system using quadrant identification signals, the signals are employed in the missile control system. The tracking pulses are not used by the missile. With such an arrangement, jamming is more difficult since the coding of the pulses by the time spacing between them is the only information desired. The quadrant signals can be thought of as *turn* commands — UP, DOWN, RIGHT, LEFT. If the missile is on the axis of the cone, all *turn* signals balance, and the missile continues its flight without correction. If the missile is displaced from the cone axis, one signal overbalances the other. The



Beam Rider Receiver Using Pulse Position Reference

missile obeys the stronger signal to turn back toward the beam axis.

Beam-rider guidance data received by the missile's antennas is detected and sent as coded pulses to the guidance receiver. After sufficient video amplification has been accomplished, the signals are decoded in suitable coincidence circuits and then filtered and transmitted to the control system as *up-down* and *right-left* error signals. These signals appear as electrical voltages proportional to the missile deviation from the beam. The control system then institutes appropriate action to correct any off-course error.

The beam-riding missile can be distinguished by the fact that the antenna or antennas for guidance are located on the aft section, or rearward looking from the missile.

It would be detrimental to the guidance system within the missile itself to pick up the echo signal. Because of the dispersion and scintillation of the returned signal there would be confusion as to the quadrant identification. The guidance antenna system of a beam rider therefore has its pickup lobes to the rear of the missile.

Because of the low traffic-handling capabilities of a beam-rider system, it is of interim use only at the present time.

We noted before that the distance over which a guidance system can function dictates the use to which it is put. Short-range systems are best used to guide missile to close targets, as in the case of aircraft interception.

index

A

B

C

G

H

I

J

K

L

R

S

AIR FORCE-UNIVERSAL PRINTING CO., ST. LOUIS, MO., 20 SEPTEMBER 1957. 5100 COPIES (580 PAGES)
Copies of this manual cannot be obtained from the printers. See Foreword page for Official distribution data.